



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification 6 : C12N 15/31, 15/00, 15/11, C07K 14/315, C07H 21/04, A61K 39/40, 39/395, G01N 33/569, G11B 5/00</p>	A1	<p>(11) International Publication Number: <b>WO 98/26072</b></p> <p>(43) International Publication Date: 18 June 1998 (18.06.98)</p>
<p>(21) International Application Number: PCT/US97/22578</p> <p>(22) International Filing Date: 9 December 1997 (09.12.97)</p> <p>(30) Priority Data: 60/036,281 13 December 1996 (13.12.96) US</p> <p>(71) Applicant (for all designated States except US): ELI LILLY AND COMPANY [US/US], Lilly Corporate Center, Indianapolis, IN 46285 (US).</p> <p>(72) Inventors; and (75) Inventors/Applicants (for US only): BALTZ, Richard, H. [US/US]; 6438 North Olney, Indianapolis, IN 46220 (US). BURGETT, Stanley, G. [US/US]; 818 Bitter Bark Lane, Indianapolis, IN 46227 (US). DEHOFF, Bradley, S. [US/US]; 7727 Maradona Drive South, Indianapolis, IN 46214 (US). HOSKINS, Jo, Ann [US/US]; 8229 Tern Court, Indianapolis, IN 46256 (US). JASKUNAS, Stanley, R., Jr. [US/US]; 16 Algonquian Drive, Natick, MA 07160 (US). MILLS, Bradley, J. [US/US]; 321 South Poplar, Morristown, IN 46161 (US). NORRIS, Franklin, H. [US/US]; 3518 South Arlington Avenue, Indianapolis, IN 46237 (US). PEERY, Robert, B. [US/US]; 372 Sycamore Street, Brownsburg, IN 46112 (US). ROCKEY, Pamela, K. [US/US]; 679 Cary</p>		<p>nary Creek Drive, Franklin, IN 46131 (US). ROSTECK, Paul, R., Jr. [US/US]; 4247 Saffron Drive, Indianapolis, IN 46237 (US). SKATRUD, Paul, L. [US/US]; 5233 Tony Drive, Greenwood, IN 46143 (US). SMITH, Michele, C. [US/US]; 4515 North Lakeridge Drive, Indianapolis, IN 46234 (US). SOLENERBERG, Patricia, J. [US/US]; 10116 Seabreeze Way, Indianapolis, IN 46256 (US). TREADWAY, Patti, J. [US/US]; 4268 West Fox Ridge Avenue, Greenwood, IN 46143 (US). YOUNG BELLIDO, Michele, L. [CA/US]; 4925 Common Vista Court, Indianapolis, IN 46220 (US).</p> <p>(74) Agents: WEBSTER, Thomas, D. et al.; Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285 (US).</p> <p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b> With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</p>
<p>(54) Title: STREPTOCOCCUS PNEUMONIAE DNA SEQUENCES</p> <p>(57) Abstract</p> <p>Nucleotide fragments of the genome of the bacterium <i>Streptococcus pneumoniae</i> are provided. Also provided are ORFs encoded by said genome.</p>		

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	YN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

-1-

## STREPTOCOCCUS PNEUMONIAE DNA SEQUENCES

5 This invention provides DNA sequences from the *Streptococcus pneumoniae* genome, and methods of use of DNA fragments originating therefrom in a variety of biological and pharmaceutical applications.

The recent emergence of widespread antibiotic  
10 resistance in common pathogenic bacterial species has justifiably alarmed the medical and research communities. Frequently these organisms are co-resistant to several different antibacterial agents. Particularly problematic has been the emergence and rapid spread of penicillin resistance  
15 in *Streptococcus pneumoniae*, which frequently causes upper respiratory tract infections. Resistance to penicillin in this organism can be due to modifications of one or more of the penicillin-binding proteins (PBPs). Combating the phenomenon of increasing resistance to antibiotic agents  
20 among pathogenic organisms such as *Streptococcus pneumoniae* will require intensified research into the fundamental molecular biology of such organisms. Greater knowledge about the molecular biology of pathogenic organisms will lead to new antibacterial agents having novel and effective actions.

25 While inroads in the development of new antibiotics and new targets for antibiotic compounds have been made with a variety of microorganisms, progress has been less apparent in *Streptococcus pneumoniae*. In part, *Streptococcus pneumoniae* presents a special case because this organism is  
30 highly recombinogenic and readily takes up exogenous DNA from its surroundings. Thus, the need for new antibacterial compounds and new targets for antibacterial therapy in *Streptococcus pneumoniae* is more acute than in other organisms.

-2-

The present invention relates to the genome of *S. pneumoniae*. The genomic information disclosed by the present invention enables: (1) preparation of molecular hybridization probes for use in PCR amplification of genes and regulatory regions, physical mapping, sequencing, mutagenesis, and mutation analysis, (2) homology comparisons with the genomes and open reading frames (ORFs) of other organisms, (3) creation of specifically mutated strains of *S. pneumoniae* wherein the mutation is targeted to any site or sites in the DNA sequence disclosed herein, (4) identification of *S. pneumoniae* promoters and other gene regulatory sequences, (5) identification of proteins/ORFs encoded by *S. pneumoniae*, (6) identification of virulence genes in *S. pneumoniae*, (7) determination of the biological function of proteins/ORFs and RNAs encoded by *S. pneumoniae*, (8) production of kits useful for determining gene function in the cell, and kits for isolating and analyzing genes that are mutated in antibiotic resistant clinical isolates of *S. pneumoniae*, (9) production of proteins and RNAs encoded by *S. pneumoniae*, (10) production of antibodies against proteins and other antigens encoded by *S. pneumoniae*, (11) methods to identify compounds that bind to proteins and RNAs encoded by *S. pneumoniae* as potential new antibiotic compounds.

In another embodiment the invention relates to substantially purified proteins encoded by the *S. pneumoniae* genome.

Table 1 summarizes the proteins and nucleic acids disclosed herein, contigs, SEQ ID NO's and predicted functions.



-3-

"Genome" refers to the full complement of chromosomal and extra-chromosomal DNA within a cell. The genome comprises the genetic blueprint for all proteins and RNAs encoded by the cell or organism.

5 "ORF" (i.e. "open reading frame") designates a region of genomic DNA beginning with a Met or other initiation codon and terminating with a translation stop codon, potentially encoding a protein product. "Partial ORF" means a portion of an ORF as disclosed herein such that the  
10 initiation codon, the stop codon, or both are not disclosed.

"DNA chip" or "Bio Chip" or "Bio DNA chip" refers to a solid matrix or support onto which is applied an array of oligonucleotides, or nucleotide sequences, or gene  
15 fragments, or genomic fragments, of *S. pneumoniae* which may further comprise a layer of *S. pneumoniae* cells suspended thereover in a semisolid medium such as agar or agarose.

"Consensus sequence" refers to an amino acid or nucleotide sequence that may suggest the biological function of a protein, DNA, or RNA molecule. Consensus sequences are  
20 identified by comparing proteins, RNAs, and gene homologs from different species.

"Contiguous fragment building" or "Contiguous fragment" or "Contig" refers to the process and result, respectively, by which a fragment of DNA is assembled from smaller  
25 constituent DNA fragments by arranging the constituent pieces in their correct order and register such that the resulting contiguous fragment accurately depicts the native DNA sequence from which the smaller fragments originated.

"Computer readable medium" includes, for example, a  
30 floppy disc, hard disc, random access memory, read only memory, and CD-ROM.

The terms "cleavage" or "restriction" of DNA refers to the catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in

-4-

the DNA (viz. sequence-specific endonucleases). The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors, and other requirements are used in the manner well known to one of ordinary skill in the art. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer or can be found in the literature.

"Diagnostics" as used herein relates to *in vitro* or *in vivo* diagnosis for disease states or biological status in mammals, preferably humans.

"Therapeutics" and "therapeutic/diagnostic combinations" means the treatment, or diagnosis and treatment, of disease states or biological status by *in vivo* administration to mammals, preferably humans, of compositions of the present invention, for example, antibodies.

"Essential genes" or "essential ORFs" or "essential proteins" refer to genomic information or the protein(s) or RNAs encoded therefrom, which, when disrupted by knockout mutation, or by other mutation, produce inviability in cells harboring said mutation.

"Non-essential genes" or "non-essential ORFs" or "non-essential proteins" refer to genomic information or the protein(s) or RNAs encoded therefrom, which, when disrupted by knockout mutation, or other mutation, do not result in inviability of cells harboring said mutation.

"Minimal gene set" refers to a genus of about 256 genes that are conserved among different bacteria such as *M. genitalium* and *H. influenzae*. The minimal gene set appears to be necessary and sufficient to sustain life. See e.g. A. Mushegian and E. Koonin, "A minimal gene set for cellular life derived by comparison of complete bacterial genomes" *Proc. Nat. Acad. Sci.* 93, 10268 - 273 (1996).

-5-

The term "fragment thereof" denotes a fragment of a nucleic acid molecule described herein, wherein said fragment comprises a region of contiguity within said nucleic acid of at least 15 base pairs. The term may also  
5 refer to a peptide of at least 5 contiguous amino acid residues of a protein disclosed herein.

The term "plasmid" refers to an extrachromosomal genetic element. The starting plasmids herein are either commercially available, publicly available on an  
10 unrestricted basis, or can be constructed from available plasmids in accordance with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

15 "Recombinant DNA cloning vector" as used herein refers to any autonomously replicating agent, including, but not limited to, plasmids and phages, comprising a DNA molecule to which one or more additional DNA segments can or have been added.

20 The term "recombinant DNA expression vector" as used herein refers to any recombinant DNA cloning vector, for example a plasmid or phage, in which a promoter and other regulatory elements are present to enable transcription of the inserted DNA.

25 The term "vector" as used herein refers to a nucleic acid compound used for introducing exogenous DNA into host cells. A vector comprises a nucleotide sequence which may encode one or more protein molecules. Plasmids, cosmids, viruses, and bacteriophages, in the natural state  
30 or which have undergone recombinant engineering, are examples of commonly used vectors.

The terms "complementary" or "complementarity" as used herein refers to the capacity of purine and pyrimidine nucleotides to associate through hydrogen bonding in double

-6-

stranded nucleic acid molecules. The following base pairs are complementary: guanine and cytosine; adenine and thymine; and adenine and uracil.

"Oligonucleotide" refers to a short polymeric nucleotide chain comprising from about 2 to 25 nucleotides.

"Isolated nucleic acid compound" refers to any RNA or DNA sequence, however constructed or synthesized, which is locationally distinct from its natural location.

A "primer" is a nucleic acid fragment which functions as an initiating substrate for enzymatic or synthetic elongation of a nucleic acid molecule.

The term "promoter" refers to a DNA sequence which directs transcription of DNA to RNA.

A "probe" as used herein is a labeled nucleic acid compound which can be used to hybridize with another nucleic acid compound.

The term "hybridization" or "hybridize" as used herein refers to the process by which a single-stranded nucleic acid molecule joins with a complementary strand through nucleotide base pairing.

"Recorded" as used herein refers to a process for storing information on a computer readable medium.

"Substantially identical" means a sequence having sufficient homology to hybridize under high stringency conditions and/or at least 90% identity at the nucleotide or amino acid sequence level to a sequence disclosed herein.

"Substantially purified" when used in reference to a protein or peptide means that the molecule has been largely, but not necessarily wholly, separated and purified from other cellular and non-cellular components. Typically a protein is substantially pure when it is at least about 60% by weight, free from other naturally occurring organic molecules. Preferably the purity is at least about 75%, more

-7-

preferably at least about 90%, and most preferably at least about 99% by weight pure.

"Selective hybridization" refers to hybridization under conditions of high stringency. Hybridization of nucleic acid molecules depends upon factors such as the degree of complementarity, stringency of hybridization conditions, and the length of hybridizing strands.

The term "stringency" relates to nucleic acid hybridization conditions. High stringency conditions disfavor non-homologous base pairing. Low stringency conditions have the opposite effect. Stringency may be altered, for example, by changes in temperature and salt concentration. Typical high stringency conditions comprise hybridizing at 50°C to 65°C in 5X SSPE and 50% formamide, and washing at 50°C to 65°C in 0.5X SSPE; typical low stringency conditions comprise hybridizing at 35°C to 37°C in 5X SSPE and 40% to 45% formamide and washing at 42°C in 1X-2X SSPE.

"SSPE" denotes a hybridization and wash solution comprising sodium chloride, sodium phosphate, and EDTA, at pH 7.4. A 20X solution of SSPE is made by dissolving 174 g of NaCl, 27.6 g of  $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ , and 7.4 g of EDTA in 800 ml of  $\text{H}_2\text{O}$ . The pH is adjusted with NaOH and the volume brought to 1 liter.

"SSC" denotes a hybridization and wash solution comprising sodium chloride and sodium citrate at pH 7. A 20X solution of SSC is made by dissolving 175 g of NaCl and 88 g of sodium citrate in 800 ml of  $\text{H}_2\text{O}$ . The volume is brought to 1 liter after adjusting the pH with 10N NaOH.

"Virulence gene" as used herein means a gene from a pathogenic organism such as *S. pneumoniae* that is required for infection and/or pathogenicity *in vivo*. Some virulence

-6-

genes are induced during infection of a host; others are expressed exclusively during in vivo infection.

The *Streptococcus pneumoniae* genome contains about 2.2 million nucleotide base pairs and comprises about 2000 to 3000 ORFs and other genes. This invention provides, among other things, contiguous fragments, genes, and proteins from the *S. pneumoniae* genome (SEQ ID NO:1 through SEQ ID NO:226).

Strain differences in *S. pneumoniae* may be associated with nucleotide sequence differences in one or more of the genomic fragments disclosed herein. Sequences that are substantially identical to the sequences disclosed herein are intended to be within the scope of the invention.

The sequence fragments disclosed herein provide a wide variety of utilities. For example, the fragments may be used to identify regions of the *S. pneumoniae* genome that are expressed as proteins (viz. transcribed into mRNA). The genomic fragments disclosed herein can also be used to examine differential expression of *S. pneumoniae* genes under diverse environmental conditions, as occurs, for example, with the expression of virulence genes during in vivo infection of a host organism. Also contemplated by the invention are: (1) preparation of molecular hybridization probes for use in physical mapping, sequencing, mutagenesis, mutation analysis, (2) homology comparisons of the sequences disclosed herein with the genomes and ORFs of other organisms, (3) creation of specifically mutated strains of *S. pneumoniae* wherein the mutation is targeted to any site in the DNA sequence disclosed herein, (4) identification of *S. pneumoniae* promoters and other gene regulatory sequences, (5) identification of proteins and RNAs encoded by *S. pneumoniae*, (6) amplification of *S. pneumoniae* genes using the PCR, and (7) production of kits for isolating and

-9-

analyzing genes that are mutated in antibiotic resistant clinical isolates of *S. pneumoniae*.

#### Genome Analysis

5 In one embodiment, the invention comprises the ORFs and fragments thereof encoded by the nucleotide sequences disclosed herein. Some of the nucleotide sequences disclosed herein encode ORFs and fragments of ORFs (Table 1). The ORFs or fragments thereof were identified by translation of the  
10 nucleic acid sequences disclosed herein. The biological function of a protein disclosed in Table 1 was determined by homology comparison with known proteins from other organisms. A number of computer programs are available to assist in homology comparisons, for example Genemark  
15 (Borodovsky and McIninch, *Computers Chem.* 17(2), 123, 1993).

#### Computer-Related Applications

The nucleotide and/or amino acid sequence information of this invention may be provided in a variety of media to  
20 facilitate use. In one embodiment the present invention comprises one or more of the sequences disclosed herein recorded on a computer readable medium. A variety of media are contemplated, for example, magnetic storage media such as floppy discs, hard disc storage, magnetic tape, and CD-  
25 ROM. A skilled artisan can readily adopt any presently known method for recording information on a computer readable medium to generate manufactures comprising the nucleotide or amino acid sequence information of the present invention. These embodiments are contemplated within the scope of this  
30 invention.

The choice of a data storage structure will generally be based on the means chosen to access the stored information. A variety of data processor programs and formats can be used to store the sequence information of the

-10-

invention on computer readable medium. For example, the sequence can be represented in a word processing text file that is formatted in commercially available software such as WordPerfect and MicroSoft Word, or it can be represented in the form of a text only file such as ASCII.

Having *S. pneumoniae* genomic sequence information in a computer readable format enables a skilled artisan to access the information for a variety of purposes. For example, computer-assisted searching algorithms may be used to identify open reading frames, and ascertain biological function based on homology to known proteins from other organisms. Suitable algorithms for sequence comparisons include BLAST (Altschul et al., *J. Mol. Biol.* 215, 403-410, 1990) and BLAZE (Brutlag et al., *Comp. Chem.* 17, 203-207 (1993)). For identification of ORFs a number of commercially available software programs are suitable, such as FRAMES (Genetic Center Group, Madison, WI).

The genomic information of this invention in computer-readable form can be manipulated further using bioinformatics to identify the biological function of proteins encoded by ORFs as well as the cellular location of said proteins. The skilled artisan will recognize several computer-assisted algorithms for this purpose, for example, PSORT which is useful for determining the likely location of a protein within a cell (See K. Nakai & M. Kanehisa. "Expert system for predicting protein localization sites in Gram-negative bacteria", *Proteins: Structure, Function, and Genetics*, 11, 95-110 (1991)).

#### Open Reading Frames and Proteins

The invention also provides proteins encoded by the *S. pneumoniae* genome in substantially purified form (See Table 1). The proteins are classified herein as (1) Hypothetical,



-11-

(2) Cell wall biosynthetic, (3) External target, or (4) Minimal gene set proteins.

Cells that carry knockout mutations in proteins of the hypothetical class are nonviable. Loss of viability suggests  
5 that these proteins may be essential for viability. Two such proteins, whose genes map to contigs m014 and m016, correspond respectively to *Haemophilus influenzae* ORFs HI1146 and HI1648. Two other hypothetical proteins, yyaF and ywbL, correspond to a GTP binding protein and  
10 transcriptional regulator, respectively.

The proteins of this invention can be used to raise antibodies. Antibodies against the hypothetical class of proteins are especially attractive. In targeting  
presumptively essential cellular functions, antibodies  
15 against "hypothetical proteins" could have therapeutic or prophylactic applications. Additionally, the "hypothetical" proteins can be used to screen for agents that bind or otherwise interact with said proteins. Such agents could lead to the identification of new antibacterial agents.

20 Proteins classified in Table 1 as cell wall biosynthetic proteins, and external target proteins, were identified by homology with known proteins. These proteins are useful for identifying agents that bind and inhibit bacterial growth. Therefore, in another embodiment of the  
25 invention, the proteins of these classifications are prepared, preferably by recombinant means as described herein, substantially purified, and used in a screen to identify compounds that bind and/or inhibit the activity of said proteins. A variety of suitable screens are  
30 contemplated for this purpose. For example, the protein(s) can be labeled by known techniques such as radiolabeling or fluorescent tagging, or by labeling with biotin/avidin; thereafter binding of a test compound to a labeled protein

-12-

can be determined by any suitable means, well known to the skilled artisan.

The proteins categorized as "minimal gene set" are homologous to a set of highly conserved proteins found in other bacteria. The minimal gene set proteins are thought to be essential for viability, and are useful targets for the development of new antibacterial compounds.

#### DNA Chips and Applications

The nucleic acids disclosed herein, or subfragments thereof, may be arrayed on any suitable solid surface, thereby constructing a "chip." DNA chip hybridizations provide greater sensitivity than do conventional hybridization means, such as Southern hybridization or Northern hybridization. DNA chips are useful for a variety of purposes, for example, in mutation and gene expression analysis, and in probing the structure, function, and expression of the genome. This aspect of the invention relates to any one or more of the DNA fragments disclosed herein, wherein said fragments are attached to a solid support (i.e. "chip" or "DNA chip" or "Bio chip"). Attachment of a nucleic acid to a support can be, but is not necessarily, accomplished by chemical or enzymatic means.

In one embodiment, DNA fragments of this invention are arrayed onto a solid support as a means for assessing gene expression in *S. pneumoniae*. The DNA fragments attached to a chip may be of any size that is suitable for hybridization to other nucleic acid molecules such as cDNAs, genomic DNAs, or RNAs. Suitably-sized DNA fragments are from 10 nucleotide residues to approximately several thousand residues. The preferred length is about 50 to 500 nucleotides.

Analysis of gene expression using the chips of this invention is assessed by hybridization of a chip to RNA samples, or cDNA samples prepared from *S. pneumoniae* grown

-13-

under any suitable conditions. Preferred samples for hybridization to a chip comprise cDNA. Methods for preparing RNA or cDNA are well known in the art.

A variety of suitable methods are known for fixing DNA fragments to solid support matrices [See e.g. D. Stimpson et al. "Real-time detection of DNA hybridization and melting on oligonucleotide arrays by using optical wave guides" *Proc. Nat. Acad. Sci.* 92, 6379 (1995)] Preferred surfaces for producing a chip are glass or polystyrene. Convenient surfaces are microscope slides, or cover slips (Corning), treated with silicon or silane to minimize non-specific binding by DNA or proteins. Also suitable for this purpose are 96-well microtiter plates.

A light-directed method may be used for attaching oligonucleotides, enabling nucleotide synthesis directly on the solid surface using photolabile 5'protected N-acyl-deoxynucleotide phosphoramidites and surface linker chemistry (See Pease et al. "Light-generated oligonucleotide arrays for rapid DNA sequence analysis" *Proc. Nat. Acad. Sci.* 91, 5022-5026, 1994). Alternatively, DNA fragments can be bound to a surface via interaction with a specific DNA binding protein. Any suitable DNA binding protein may be used, for example bacteriophage DNA binding proteins, Adenovirus binding protein, the *E. coli* lac-repressor protein, or  $\lambda$ -repressor protein. DNA binding proteins are attached to the surface of a chip by covalent chemical binding, essentially as described in U.S. Patent 5,561,071, the entire contents of which is incorporated by reference. The latter method requires that DNA fragments contain a recognition sequence that enables binding by the DNA binding protein. Specific sequences for a number of DNA binding proteins are known. Methods for incorporating specific binding sequences into the genomic DNA fragments disclosed herein are well known in the cloning arts.

-14-

DNA chip technology enables monitoring *S. pneumoniae* gene expression on a genome-wide level. This feature of the invention is particularly attractive for identifying (1) genes that are expressed or not expressed during the life cycle or infection cycle of *S. pneumoniae*, and (2) changes in gene expression that correlate with environmental change.

For example, virulence genes in *S. pneumoniae* can be identified by the DNA chip method disclosed herein.

Identification of virulence genes in *S. pneumoniae* will provide new targets for developing novel antibiotics. For this aspect of the invention any suitable encapsulated strain of *S. pneumoniae* is introduced into a mouse, for example, by intraperitoneal injection, or by introduction directly into the lungs, or by any other suitable method. Approximately 2 days after infection a peripheral blood titre level is reached of about  $10^8$  *S. pneumoniae* cells/ml. Cells recovered from peripheral blood, or other suitable tissue, are used in identifying virulence genes. For this purpose, cDNAs are prepared from cells recovered from an *in vivo* infection and from cells grown *in vitro*. After labeling, the cDNAs are hybridized against the DNA chip(s) disclosed herein. Genomic fragments that hybridize to the *in vivo* probe but not to the *in vitro* probe identify candidate virulence genes.

Also contemplated by this aspect of the invention is a method for analyzing gene expression in *S. pneumoniae* cells grown or harvested from any desirable *in vitro* or *in vivo* environment, wherein said environment may include compounds whose effects on gene expression are to be determined.

In another embodiment, the present invention relates to a DNA bio-chip, useful for correlating DNA sequence with biological function. The bio-chip comprises an array of the genomic DNA fragments disclosed herein, or portions thereof, attached to the surface of any suitable solid support

-15-

material. The bio chip further comprises a layer of competent *S. pneumoniae* cells suspended over the DNA array in any suitable semi-solid medium such as agar or agarose. The cells suspended on the bio chip comprise known or  
5 unknown mutant strains, or they may be wild-type cells. The cell layer is in contact with the DNA matrix such that DNA on the chip can be taken up by the cells.

The bio-chip is useful for several purposes. For example, the bio-chip can be used to localize an unknown  
10 mutation to a specific region of the genome by complementation. The bio-chip enables correlating a phenotype with a genetic locus. For example, mutant cells harboring one or more mutations and having at least one screenable or selectable phenotype can be applied to a bio  
15 chip and subjected to an environment that allows for selection, or for screening by complementation. If said phenotype is the result of a chromosomal mutation or mutations that map to a genomic fragment present on the chip, DNA uptake by the cells and repair of the mutation by  
20 recombination will be identifiable by a suitable screen or selection.

In a preferred embodiment, the bio-chip is overlaid with competent *S. pneumoniae* cells. Methods for preparing competent cells are known (See e.g. LeBlanc et.al. *Plasmid*  
25 28, 130-145, 1992; Pozzi et al. *J. Bacteriol.* 178, 6087-6090, 1996).

Other embodiments of this aspect of the invention are contemplated. For example the genomic fragments disclosed herein could be prepared and dispensed into individual wells  
30 of a 96-well micro titre plate. Competent *S. pneumoniae* cells could then be added to the wells under conditions suitable for DNA uptake followed by plating onto any suitable selection or screening medium, for example an agar

-16-

plate containing suitable growth and/or selection/screening components.

#### Diagnostic Kits and Assays

- 5       The present invention further relates to kits and assays that can be used for rapid and efficient detection of *S. pneumoniae* cells. Also contemplated are kits for detecting mutations carried by *S. pneumoniae* cells. Kits of this nature are particularly attractive in the clinical  
10       environment where knowledge about the identity of a pathogen and/or of the basis for resistance to antibiotic treatments is essential for effective medical treatment. In the long term, knowledge of the mutations that lead to resistance will enable the design of new antibacterial agents.
- 15       A kit for detecting *S. pneumoniae* cells can be based on antibody recognition of *S. pneumoniae* specific antigens or epitopes, or by nucleic acid hybridization techniques for the detection of *S. pneumoniae* specific nucleic acid molecules.
- 20       A variety of embodiments are contemplated in this aspect of the invention. In one embodiment a kit is provided for detecting mutations in drug-resistant *S. pneumoniae*. For this purpose, DNA is prepared from a resistant isolate and from a wild-type strain. In a preferred embodiment, the  
25       polymerase chain reaction (i.e. PCR) is used to amplify DNA samples representing any one or all of the genomic fragments disclosed herein. The amplified DNAs from the mutant and wild-type cells are hybridized to a DNA chip having fixed thereon any one or more of the genomic fragments disclosed  
30       herein. Amplified DNA samples from the mutant and wild-type strain are labeled by any suitable means, for example using radioisotopes or fluorescent labeling. Hybridization of the amplified DNAs to the chip under conditions that can discriminate single or multiple base pair mismatches enables

-17-

the detection of differences between the mutant and wild-type samples. This method identifies a specific fragment of the genome that is altered in the mutant strain. The specific mutation can be determined by conventional DNA  
5 sequence analysis.

This aspect of the invention also relates to the detection of *S. pneumoniae* proteins in a sample using antibody molecules raised against any suitable ORF disclosed herein. Antibody detection methods are well known to those  
10 skilled in the art including, for example, a variety of radioimmunological assays. (See e.g. P. Tijssen, Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands, 1985).

15 Test samples suitable for use in this aspect of the invention include but are not limited to biological fluids such as sputum, blood, serum, plasma, urine, and to biopsy samples.

Skilled artisans will recognize that the disclosed  
20 method and reagents can be readily incorporated into a kit. For example, a kit would contain one or more receptacles comprising one or more of the following: PCR reagents, DNA chip reagents, labeling reagents, assorted buffers, and/or antibodies.

#### 25 Production of Antibodies

The proteins of this invention and fragments thereof may be used in the production of antibodies. The term "antibody" as used herein describes antibodies,  
30 fragments of antibodies (such as, but not limited, to Fab, Fab', Fab<sub>2</sub>', and Fv fragments), and chimeric, humanized, veneered, resurfaced, or CDR-grafted antibodies capable of binding antigens of a similar nature as the parent antibody

-18-

molecule from which they are derived. The instant invention also encompasses single chain polypeptide binding molecules.

The production of antibodies, both monoclonal and polyclonal, in animals is well known in the art. See, e.g.,  
5 C. Milstein, Handbook of Experimental Immunology, (Blackwell Scientific Pub., 1986); J. Goding, Monoclonal Antibodies: Principles and Practice, (Academic Press, 1983). For the production of monoclonal antibodies the process begins with  
10 injecting a mouse, or other suitable animal, with an immunogen. The mouse is subsequently sacrificed and cells taken from its spleen are fused with myeloma cells, resulting in a hybridoma that can be cultured *in vitro*. Hybridomas are screened for clones that secrete a single  
15 antibody species, specific for the immunogen.

Chimeric antibodies, described in U.S. Patent No. 4,816,567, herein incorporated by reference, teaches methods and vectors for preparing chimeric antibodies. An  
20 alternative approach is provided in U.S. Patent No. 4,816,397, the entire contents of which is herein incorporated by reference. This patent teaches co-expression of heavy and light chains in the same host cell.

The method taught in U.S. Patent 4,816,397 has been further refined in European Patent Publication No. 0 239 400. The teachings of this publication are preferred for  
25 engineering monoclonal antibodies. In this technology the complementarity determining regions (CDRs) of a human antibody are replaced with the CDRs of a murine monoclonal antibody, thereby converting the specificity of the human antibody to the specificity of the murine antibody.

30 Single chain antibodies and libraries thereof provide yet another means for genetically engineering antibody molecules. (See, e.g. R.E. Bird, et al., *Science* 242:423-426 (1988); PCT Publication Nos. WO 88/01649, WO 90/14430, and WO 91/10737. Single chain antibody technology



-19-

involves covalently joining the binding regions of heavy and light chains thereby generating a single polypeptide chain having the binding specificity of an intact antibody molecule.

5       The antibodies contemplated by the present invention are useful in diagnostics, therapeutics, or in diagnostic/therapeutic combinations.

      The proteins of this invention, or suitable fragments thereof, can be used to generate polyclonal or monoclonal  
10       antibodies, and various inter-species hybrids, or humanized antibodies, or antibody fragments, or single-chain antibodies. The techniques for producing antibodies are well known to skilled artisans. (See e.g. A.M. Campbell,  
      Monoclonal Antibody Technology: Laboratory Techniques in  
15       Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam (1984); Kohler and Milstein, *Nature* 256, 495-497 (1975); Monoclonal Antibodies: Principles & Applications Ed. J.R.Birch & E.S. Lennox, Wiley-Liss, 1995.

      A protein or peptide to be used as an immunogen may be  
20       administered in an adjuvant by subcutaneous or intraperitoneal injection into, for example, a mouse or a rabbit. For the production of monoclonal antibodies, spleen cells from immunized animals are removed, fused with myeloma cells, such as SP2/0-Agl4 cells, and allowed to become  
25       monoclonal antibody producing hybridoma cells in the manner known to the skilled artisan. Hybridomas that secrete the desired antibody molecule can be screened by a variety of well known methods, for example ELISA assay, western blot analysis, or radioimmunoassay (Lutz et al. *Exp. Cell Res.*  
30       175, 109-124 (1988); Monoclonal Antibodies: Principles & Applications Ed. J.R.Birch & E.S. Lennox, Wiley-Liss, 1995).

      For some applications it is desirable to have an antibody labeled in some fashion. Procedures for labeling antibody molecules with radioisotopes, affinity labels, such

-20-

as biotin or avidin, enzymatic labels, for example horseradish peroxidase, and fluorescent labels such as FITC or rhodamine, are widely known (See e.g. Enzyme-Mediated Immunoassay, Ed. T. Ngo, H. Lenhoff, Plenum Press 1985;

- 5 Principles of Immunology and Immunodiagnostics, R.M. Aloisi, Lea & Febiger, 1988).

Labeled antibodies are useful for a variety of diagnostic applications. In one embodiment, the present invention relates to the use of labeled antibodies to detect  
10 the presence of *S. pneumoniae* cells and proteins. Also contemplated are applications that use antibodies, preferably single chain antibodies, directed against a *S. pneumoniae* protein. Proteins identified as "external targets" are preferred for the generation of single chain  
15 antibodies. Single chain antibody libraries directed against *S. pneumoniae* surface proteins and cell wall proteins can be produced by applying the phage display technique to crude membrane preparations. Antibodies that recognize and bind to external target proteins and/or cell wall proteins could be  
20 used as therapeutic agents to inhibit the growth of *S. pneumoniae*. Alternatively, the antibodies could be used in a screen to identify potential inhibitors of an external target protein. For example, in a competitive displacement assay, an antibody or compound to be tested is labeled by  
25 any suitable method. Competitive displacement of an antibody from an antibody-antigen complex by a test compound provides a means to identify new antibacterial compounds.

#### Protein Production Methods

- 30 The present invention relates further to substantially purified proteins encoded by the ORFs disclosed herein (SEQ ID NO:87 through SEQ ID NO:228).

Skilled artisans will recognize that proteins can be synthesized by different methods, for example, chemical

-21-

methods or recombinant methods, as described in U.S. Patent 4,617,149, hereby incorporated by reference.

The principles of solid phase chemical synthesis of polypeptides are well known in the art and may be found in general texts relating to this area. See, e.g., H. Dugas and C. Penney, Bioorganic Chemistry (1981) Springer-Verlag, New York, 54-92. Peptides may be synthesized by solid-phase methodology utilizing an Applied Biosystems 430A peptide synthesizer (Applied Biosystems, Foster City, CA) and synthesis cycles supplied by Applied Biosystems. Protected amino acids, such as t-butoxycarbonyl-protected amino acids, and other reagents are commercially available from many chemical supply houses.

The proteins and peptides of the present invention can also be made by recombinant DNA methods. Recombinant methods are preferred if a high yield is desired. Recombinant methods involve expressing a cloned ORF/gene in a suitable host cell. A gene is introduced into a host cell by any suitable means, well known to those skilled in the art. While chromosomal integration of a cloned gene is within the scope of the present invention, it is preferred that a cloned gene be maintained extra-chromosomally, as part of a vector wherein the gene is in operable-linkage to a constitutive or inducible promoter.

Recombinant methods are also useful in overproducing a membrane-bound or membrane-associated protein. In some cases, membranes prepared from recombinant cells that overexpress such proteins provide an enriched source of the protein. Such membranes are useful for evaluating the function of the protein and/or for evaluating inhibitors of the protein.

-22-

# Expressing Recombinant Proteins in Procarvotic and Eucaryotic Host Cells

Procarvytes are generally used for cloning DNA sequences and for constructing vectors. For example, the  
 5 *Escherichia coli* K12 strain 294 (ATCC No. 31446) is particularly useful for expression of foreign proteins. Other strains of *E. coli*, bacilli such as *Bacillus subtilis*, enterobacteriaceae such as *Salmonella typhimurium* or  
 10 *Serratia marcescans*, various *Pseudomonas* species may also be employed as host cells in cloning and expressing the recombinant proteins of this invention. Also contemplated are various strains of *Streptococcus* and *Streptocomyces*.

For effective expression of a recombinant protein a gene or ORF may be linked to a known promoter sequence.  
 15 Suitable bacterial promoters include b<sup>-</sup>-lactamase [e.g. vector pGX2907, ATCC 39344, contains a replicon and b<sup>-</sup>-lactamase gene], lactose systems [Chang et al., *Nature* (London), 275:615 (1978); Goeddel et al., *Nature* (London), 281:544 (1979)], alkaline phosphatase, and the tryptophan  
 20 (trp) promoter system [vector pATH1 (ATCC 37695)] designed for the expression of a trpE fusion protein. Hybrid promoters such as the tac promoter (isolatable from plasmid pDR540, ATCC-37282) are also suitable. Promoters for use in bacterial systems also will contain a Shine-Dalgarno  
 25 sequence operably linked to the DNA encoding the desired polypeptides. These examples are illustrative rather than limiting.

A variety of mammalian cell systems and yeasts are also suitable host cells. The yeast *Saccharomyces*  
 30 *cerevisiae* is a commonly used eucaryotic microorganism. Other yeasts such as *Kluyveromyces lactis* are also suitable. For expression of recombinant genes in *Saccharomyces*, the plasmid YRp7 (ATCC-40053), for example, may be used. See, e.g., L. Stinchcomb, et al., *Nature*, 282:39 (1979); J.

-23-

Kingsman et al., *Gene*, 7:141 (1979); S. Tschemper et al., *Gene*, 10:157 (1980). Plasmid YRp7 contains the TRP1 gene that provides a selectable marker in a trp1 mutant.

5 Purification of Recombinantly-Produced Protein

An expression vector carrying an ORF of the present invention is transformed or transfected into a suitable host cell using standard methods. Cells which contain the vector are propagated under conditions suitable for expression of the encoded protein. If the gene is under the control of an inducible promoter then suitable growth conditions would incorporate the appropriate inducer. The recombinantly-produced protein may be purified from cellular extracts of transformed cells by any suitable means.

15 In a preferred process for protein purification a gene/ORF is modified at the 5' end, or some other position, to incorporate a plurality of histidine residues at the amino terminus of the encoded protein. The "histidine tag" produced thereby enables a single-step protein purification method referred to as "immobilized metal ion affinity chromatography" (IMAC), essentially as described in U.S. Patent 4,569,794, hereby incorporated by reference. The IMAC method enables rapid isolation of substantially pure protein starting from a crude cellular extract.

25 As skilled artisans will recognize, the proteins of the invention can be encoded by a multitude of different nucleic acid sequences owing to the degeneracy of the genetic code. The present invention further comprises these alternate nucleic acid sequences.

30 The ribonucleic acid compounds of the present invention may be prepared using the polynucleotide synthetic methods discussed *supra*, or they may be prepared enzymatically using RNA polymerase to transcribe a DNA template.

-24-

The most preferred systems for preparing the ribonucleic acids of the present invention employ the RNA polymerase from the bacteriophage T7 or the bacteriophage SP6. These RNA polymerases are highly specific, requiring the insertion of bacteriophage-specific sequences at the 5' end of the template to be transcribed. See, J. Sambrook, et al., *supra*, at 18.82-18.84.

This invention also provides nucleic acids, RNA or DNA, which are complementary to the sequences disclosed herein.

The present invention also provides probes and primers useful for a variety of molecular biology techniques including, for example, hybridization screens of genomic or subgenomic libraries, detection and quantification of mRNA species as a means to analyzing gene expression, and amplification of any region of the *Streptococcus pneumoniae* genome disclosed by the sequences herein. A nucleic acid compound is provided comprising any of the sequences disclosed herein, or a complementary sequence thereof, or a fragment thereof, which is at least 15 base pairs in length, and which will hybridize selectively to *Streptococcus pneumoniae* DNA or mRNA. Preferably, the 15 or more base pair compound is DNA. A probe or primer length of at least 15 base pairs is dictated by theoretical and practical considerations. See e.g. B. Wallace and G. Miyada, "Oligonucleotide Probes for the Screening of Recombinant DNA Libraries," In Methods in Enzymology, Vol. 152, 432-442, Academic Press (1987).

The probes and primers of this invention can be prepared by methods well known to those skilled in the art (See e.g. Sambrook et al. *supra*). In a most preferred embodiment these probes and primers are synthesized by the polymerase chain reaction (PCR).

-25-

The present invention also relates to recombinant DNA cloning vectors and expression vectors comprising the nucleic acids of the present invention. Preferred nucleic acid vectors are those which comprise DNA. The skilled artisan understands that choosing the most appropriate cloning vector or expression vector depends on a number of factors including the availability of restriction enzyme sites, the type of host cell into which the vector is to be transfected or transformed, the purpose of the transfection or transformation (e.g., stable transformation as an extrachromosomal element, or integration into the host chromosome), the presence or absence of readily assayable or selectable markers (e.g., antibiotic resistance and metabolic markers of one type and another), and the number of gene copies desired in the host cell.

Vectors suitable to carry the nucleic acids of the present invention comprise RNA viruses, DNA viruses, lytic bacteriophages, lysogenic bacteriophages, stable bacteriophages, plasmids, viroids, and the like. The most preferred vectors are plasmids.

Host cells harboring the nucleic acids disclosed herein are also provided by the present invention. A preferred host is *E. coli* which has been transfected or transformed with a vector that comprises a nucleic acid of the present invention.

The present invention also provides a method for constructing a recombinant host cell capable of expressing an ORF disclosed herein, said method comprising transforming or otherwise introducing into a host cell a recombinant DNA vector that comprises an isolated DNA sequence which encodes said ORF. The preferred host cell is any strain of *E. coli* which can accommodate high level expression of an exogenously introduced gene. Transformed host cells are cultured under conditions well known to skilled artisans such that said ORF

-26-

is expressed, thereby producing the encoded protein in the recombinant host cell.

For the purpose of discovering new inhibitors of cell wall biosynthesis, it would be desirable to determine agents that inhibit enzymes required for synthesis of the cell wall and/or agents that interact with membrane proteins. A method for identifying compounds that interact with such enzymes and membrane proteins comprises contacting said proteins with a test compound and monitoring an interaction and/or inhibition by any suitable means.

The instant invention provides a screening system for compounds that interact with membrane proteins of this invention, said screening system comprising the steps of:

- a) preparing a membrane protein, or membranes enriched in said protein;
- b) exposing the protein source of (a) to a test compound; and
- c) quantifying the interaction of said protein with said compound by any suitable means.

The screening method of this invention may be adapted to automated procedures such as a PANDEX® (Baxter-Dade Diagnostics) system, allowing for efficient high-volume screening of compounds.

In a typical screening protocol, a protein to be tested is prepared as described herein, preferably using recombinant DNA technology. A test compound is introduced into a reaction vessel containing said protein. The reaction/interaction of said protein and said compound is monitored by any suitable means. For example, a radioactively-labeled or chemically-labeled compound or



-27-

protein may be used. Specific association between a test compound and protein is monitored by any suitable means.

The following examples more fully describe the present invention. Those skilled in the art will recognize that the particular reagents, equipment, and procedures described are merely illustrative and are not intended to limit the present invention in any manner.

## EXAMPLE 1

10     Vector for Expressing *S. pneumoniae* ORF in a Host Cell

          An expression vector suitable for expressing a *S. pneumoniae* gene or fragment thereof in a variety of procaryotic host cells, such as *E. coli*, is easily made. A suitable parent vector contains an origin of replication (Ori), a marker for selecting transformants, for example, an ampicillin resistance gene (Amp), and further comprises suitable transcriptional and translational signals, for example, the T7 promoter and T7 terminator sequences, in operable-linkage to a *S. pneumoniae* coding region. For example, pET11A (obtained from Novogen, Madison WI) is linearized by restriction with endonucleases NdeI and BamHI. Linearized pET11A is ligated to a DNA fragment bearing NdeI and BamHI sticky ends and comprising a coding region for a *S. pneumoniae* ORF.

25           The ORF used in this construction may be modified at the 5' end (amino terminus of encoded protein or peptide) to simplify purification of the encoded protein or peptide. For this purpose, an oligonucleotide encoding 8 histidine residues is inserted after the transcriptional and translational start sites. Placement of the histidine residues at the amino terminus of the encoded protein enables the IMAC one-step protein purification procedure.

## Example2

-26-

Recombinant Expression and Purification of a Protein Encoded  
by a *S. pneumoniae* ORF

- An expression vector that carries an ORF from the *S. pneumoniae* genome, as disclosed in Example 1, and which
- 5 ORF is operably-linked to an expression promoter, is transformed into *E. coli* BL21 (DE3) (*hsdS gal lacIts857 ind1Sam7nin5lacUV5-T7gene 1*) using standard methods. Transformants, selected for resistance to ampicillin, are chosen at random and tested for the presence of the vector
- 10 by agarose gel electrophoresis using quick plasmid preparations. Colonies that contain the vector are grown in L broth and the protein produced by the vector-borne ORF is purified by IMAC, essentially as described in US Patent 4,569,794.
- 15 Briefly, the IMAC column is prepared as follows. A metal-free chelating resin (e.g. Sepharose 6B IDA, Pharmacia) is washed in distilled water to remove preservatives and then infused with a suitable metal ion [e.g. Ni(II), Co(II), or Cu(II)] by adding a 50mM metal
- 20 chloride or metal sulfate aqueous solution until about 75% of the interstitial spaces of the resin are saturated with colored metal ion. The column is then ready to receive a crude cellular extract containing the recombinant protein product.
- 25 Unbound proteins and other materials are removed by washing the column with any suitable buffer, pH 7.5. Bound protein is eluted in any suitable buffer at pH 4.3, or preferably with an imidazole-containing buffer at pH 7.5.

30 Example 3

DNA Chip Production

Any one or more of the *S. pneumoniae* genome DNA fragments disclosed herein, or fragments thereof, are arrayed onto a solid support. It is preferred that fragments be in

-29-

the size range of 14 base pairs to 500 base pairs. The DNA samples are most conveniently synthesized by PCR using standard methods to amplify regions disclosed by the genomic sequences herein. The method of Schena et al. is used to spot about 1 ng to 10 ng of a DNA sample onto glass microscope slides that have been treated with poly-L-lysine (M. Schena et al. "Quantitative monitoring of gene expression patterns with a complementary DNA microarray" Science, 270, 467-470, 1995). After spotting DNA samples onto the chip and air-drying, the chips are rehydrated by incubation for about 2 hours in a humid chamber. Chips are then placed at 100° C for 1 minute, rinsed in 0.1% SDS, and treated with 0.05% succinic anhydride in 50% 1-methyl-2-pyrrolidinone and 50% boric acid.

## Example 4

S. pneumoniae Gene Expression Analysis using DNA Chips

RNA prepared from cells grown under any desirable conditions is used to prime cDNA synthesis by reverse transcription, using methods well known to the skilled artisan (See e.g. Molecular Cloning, 2d Ed. J.Sambrook, E. Fritsch, T. Maniatis, 1989). For example, total RNA of strain R6 is prepared according to the method of Logeman et.al., (Analytical Biochemistry, 1987, 163, 16-20) using guanidine hydrochloride. After ethanol precipitation, the total RNA is dissolved in a buffered solution such as Tris-EDTA (TE). Complementary DNA's are synthesized with the aid of the StrataScript RT-PCR kit (Stratagene, Inc.) in accordance with the supplier's recommendations (See Schena et al. Id.). Briefly, a 50 ul reaction contains about 0.1 ug/ul of RNA. First strand synthesis is primed using random primers, 1X first strand buffer, 0.03 U/ul ribonuclease block, 500 uM dATP, 500 uM dTTP, 40 uM dGTP, 40 uM fluorescein-12-dCTP (New England Nuclear), and 0.03 U/ul

WO 98/26072

PCT/US97/22578

-30-

reverse transcriptase. Reactions are incubated for 60 minutes at 37° C, precipitated with ethanol, and resuspended in 10 ul TE pH 8. Samples are heated for 3 minutes at 94° C and chilled on ice. The RNA is degraded by adding 0.25 ul of  
5 10 N NaOH, followed by a 10 minute incubation at 37° C. The samples are neutralized with 2.5 ul of 1M Tris-HCl, pH 8 and 0.25 ul of 10 N HCl. After ethanol precipitation, the nucleic acid pellet is washed and dried *in vacuo*.

Prior to hybridization, DNA chips prepared as in Example  
10 3 are denatured by heating to 90°C for 2 minutes. Hybridization reactions contain about 1 ul of fluorescently-labeled cDNA, and 1 ul of hybridization buffer (10x SSC and 0.2% SDS). Probe mixtures are transferred to the surface of the chip, covered with a cover slip, and incubated for 18  
15 hours at 65° C. Chips are washed 5 minutes at room temperature in 1X SSC, 0.1% SDS, then for 10 minutes at room temperature in 0.1X SSC, 0.1% SDS. After hybridization, chips are scanned with a laser-scanning device.

-31-

## Example 5

A DNA Bio Chip for mutation analysis

Duplicate DNA chips are prepared as in Example 3. Each chip is overlaid with *S. pneumoniae* cells in a semi-solid medium, wherein said cells carry a temperature-sensitive (ts) mutation in a gene required for autolytic activity (Lyt<sup>-</sup>). This mutation leads to resistance to lysis at 37° C, but sensitivity to lytic treatments at 30° C.

*S. pneumoniae* strain cwl is resistant to lysis by detergent and penicillin when grown at 37° C, but remains sensitive when grown at 30° C (cwl is derived from strain R6; See P. Garcia et al. "Mutants of Streptococcus pneumoniae that contain a temperature-sensitive autolysin" *J. Gen. Microbiol.* 132, 1401-05, 1986). Strain cwl is grown at 30° C and competent cells are prepared according to any suitable method (e.g. LeBlanc et.al. *Plasmid* 28, 130-145, 1992; Pozzi et al. *J. Bacteriol.* 178, 6087-6090, 1996). Competent cwl cells are harvested by centrifugation and resuspended at about 10<sup>5</sup> cells per ml in 1% melted agar supplemented with 0.1% (w/v) yeast extract (Difco) and containing 1% to 2% Triton X-100. Approximately 100 ul to 500 ul of the cell mixture is deposited per square centimeter onto the bio chip by pipetting onto the chip surface. After solidification of the agar layer, one of the bio-chips is incubated at 37° C and the other at 30° C. Cells that take up a complementing genomic DNA fragment from the chip surface will be lysed at both 30° C and 37° C, while non-complemented cells are lysed only at 30° C. Cells that are complemented by the bio-chip are recognizable by this phenotypic difference and can be further purified by well known methods.

TABLE 1  
S. pneumoniae Genes and  
proteins

Contig ID	Associated ORF(s)	Nucleotide SEQ ID NO (Amino Acid SEQ ID NO)	Gene category	Comments
j001	yyaF	1 (123)	hypothetical	yyaF= GTP-binding protein
j002	obg	2 (124)	external	obg=Grp-binding protein
j003	lspA; ywbI	3 (126) 3 (125)	cell wall synth; hypothetical	lsp=prolipo-protein signal peptidase ywbI=hypothetical transcriptional regulator
j004	pstA	4 (127)	hypothetical external	pstA=high-affinity phosphate-specific transport
j005	pilB	5 (128)	external	pilB=fimbrial transcription
j008	pgsA	6, 111	external	regulation repressor: pgsA=phosphatidylglycerophosphate synthase
j009	potC; potF	7 (130) 7 (131)	external	potC, potF=spermidine/potassium transport system permease
j010	pbuX; uraA(pyrP)	8, 9 (132) (133)	external	pbuX=xanthine permease; uraA=uracil permease
j012	nanB/A	10 (134)	external	nanB/A=neuraminidase

j013	spoIIIEB	11 (135)	external	spoIIIEB=sporulation protein
j014	licC	12 (136)	external	licC=OM lipopolysaccharide phase variation protein
m001	unk	13 (137)	hypothet ical	
m002	unk	14 (138)	hypothet ical	
m003	unk	15 (139)	hypothet ical	
m004	unk	16, 95 (140)	hypothet ical	
m005	unk	17 (141)	hypothet ical	
m007	murZ	18, 103	hypothet ical	
m008	unk	19 (142)	hypothet ical	
m011	unk, unk unk unk	20 (143) 20 (144) 20 (145) 20 (146)	hypothet ical hypothet ical	
m012	abctran, aminoatrans, glutrans unk (H11146)	21 (147), 22 (148), 23(149), 24 (150) 25 (151) 26, 121	external	transport proteins
m014			hypothet ical	

-34-

m015	ftsH hprt	27, 117 (153)	external	putative ATPase involved in membrane functions
m016	unk, unk (H11648)	28, 109 (154, 155)	hypothet ical	
m017	nusG	29 (156)	min set w/o euk match	transcription elongation factor
m018	unk	30 (157)	hypothet ical	
m019	era	31, 91	external	membrane associated GTP binding protein
m021	yhgF	32 (158)	external	hypothetical
m022	rpoA, secY KAD	33 (158) 115 (159)	external min set w/o euk match	rpoA encodes alpha subunit of RNA polymerase, secY is a translocator of proteins
m023	pyrG	34 (161)		(multispanning membrane protein); KAD = adenylate cyclase
m024	amiF	35 (162)	external	CTP synthetase
m027	unk	36 (163)	external	amidase
m028	spoT	37 (164)	external	guanosine 5'- diphosphate, 3'- diphosphate pyrophosphatase



m030	may	36, 99	cell wall syn/memb rane int. cell wall syn/memb rane int.	UDP-N acetylmuramoyl- pentapeptide:undecapre- nyl_P14 phosphatase
m031	rodA	39, 113	cell wall syn/memb rane int.	
m034	murB	40 (165)	cell wall syn/memb rane int.	UDP-N- acetylglucosaminyl-3- enolpyruvate reductase
m036	ftsY	41, 119	cell wall syn/memb rane int.	function in cell division
p001	nrfE; lacRrepressor nrdF	42 (166) (167) (168)	minimal set	nrfE: Nitrite reduction, formate dependent, nrdF: nucleotide reductase
p002	ACPSyn; fabB; ACPtrans; pepN; unk ligase	43, 44, 45 (172) (170) (171) (173) (169) (174)	external	fabB: fatty acid biosynthesis, pepN: aminopeptidase N
p003	trigger factor	46, 47, 97	minimal set	ligase
p004	trigger factor	48 (176)	minimal	trigger factor

p005	ptsI	49, 50 (177, 178)	set external	PRP-protein PTS enzyme I
p006	ftsZ	51, 105	external	filamentous temperature sensitivity, cell septation
p007	unk; unk; ftsE; unk	52 (180) (181) (179) (182)	Cell wall synth.	ftsE: filamentous temperature sensitivity, ATP- binding protein
p008	glpF;	53 (183)	external	glpF: glycerol facilitator
p009	grpE;	54, 107	external	grpE: nucleotide exchange factor
p010	unk; unk unk unk unk unk	55 (186) (187) (188) (189) (190) (191) (192)	cell wall synth (murI region)	
p012	topA	56 (193)	minimal set	DNA topoisomerase I
p013	deoD	57 (194)	minimal set	purine-nucleoside phosphorylase
p014	rfbX; lhc-1 operon	58, (195) 59, 60 (196-198)	cell wall synth.	rfbX: hydrophobic protein involved in O- antigen assembly

	ATPase;	61, 62, 63, 64 (199-202)		
p016			external	cation transporting ATPase
p017	livH; brac	65 (204) 65 (203)	external	livH: leucine, isoleucine and valine transport
p018	rimI; gcp	66 (205) 93	external	rimI: modification of 30S ribosomal subunit, gcp: essential gene of unknown function
Punk01	unk;	67 (206-207)	hypothetical	
Punk02	unk	68 (208)	hypothetical	
Punk03	unk	69 (209)	hypothetical	
Punk04	unk	70 (210)	hypothetical	
Punk05	unk;	71 (211-212)	hypothetical	
Punk06	unk;	72 (213-214)	hypothetical	
Punk08	unk	73 (215)	hypothetical	
Punk09	unk;	74 (216-217)	hypothetical	
Punk10	unk	75 (218)	hypothetical	
Punk12	unk	76 (219)	hypothetical	
s001	AtpF	77 (220)	minimal set	ATP synthase B chain
s002	DnaG	78, 89	minimal set	DNA primase

s003	EL-G: polC	79 (222) (221)	minimal set	elongation factor G; DNA polymerase III
s004	FAD	80 (223)	minimal set	FAD synthetase
s005	SmpB	81 (224)	minimal set	regulatory protein LPS-heptosyl-2- transferase, small prot.
s006	DnaA	82 (225)	minimal set	chromosomal replication initiator
s008	DnaB	83, 87	minimal set	replicative DNA helicase
s009	LgtP	84 (226)	minimal set	prolipoprotein diacylglyceryl transferase
s010	UDPgalactoseEpime rase; glycosyl transferase	85 (227) 86 (228)	Cell wall synth	UDPgalactose Epimerase
	dpj-acps	101		

-39-

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- 5 (i) APPLICANT: Baltz, Richard H.  
Burgett, Stanley G.  
DeHoff, Bradley S.  
10 Jaskunas Jr., Stanley R.  
Mills, Bradley J.  
Norris, Franklin H.  
Peery, Robert B.  
Rosteck Jr., Paul R.  
15 Skatrud, Paul L.  
Smith, Michele C.  
Rockey, Pamela K.  
Young-Bellido, Michele
- (ii) TITLE OF INVENTION: Streptococcus Pneumoniae DNA Sequences
- 20 (iii) NUMBER OF SEQUENCES: 122
- (iv) CORRESPONDENCE ADDRESS:  
25 (A) ADDRESSEE: Eli Lilly and Company  
(B) STREET: Lilly Corporate Center  
(C) CITY: Indianapolis  
(D) STATE: Indiana  
(E) COUNTRY: U.S.  
(F) ZIP: 46285
- 30 (v) COMPUTER READABLE FORM:  
(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
35 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:  
(A) APPLICATION NUMBER:  
40 (B) FILING DATE:  
(C) CLASSIFICATION:
- (vii) ATTORNEY/AGENT INFORMATION:  
(A) NAME: Webster, Thomas D.  
45 (B) REGISTRATION NUMBER: 39,872  
(C) REFERENCE/DOCKET NUMBER: X-11162
- (ix) TELECOMMUNICATION INFORMATION:  
(A) TELEPHONE: 317-276-3334
- 50 (2) INFORMATION FOR SEQ ID NO:1:
- (i) SEQUENCE CHARACTERISTICS:  
55 (A) LENGTH: 1267 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 60 (iii) HYPOTHETICAL: NO

- 40 -

(iv) ANTI-SENSE: NO

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

	ATGGTGAAG TTCCAGATGA ACGCCTACAA AAATACTAGT AAATGATAAC TCCTAAAAAG	60
10	ACAGTTCCCA CAACATTGTA ATTTACAGAT ATTGCAGGGA TTGTAAAAGG AGCTTCAAAA	120
	GGAGAAGGGC TAGGGAATAA ATTCTTGGCC AATATTCTGT AAGTAGATGC GATTGTTAC	180
15	GTAGTTCGTG CTTTGTATGA TGAAAATGTA ATGCGCGAGC AAGGACGTGA AGACGCCCTT	240
	GTAGATCCAC TTGCAGATAT TGATACAATT AATCTGGAAT TAATTCTTGC TGACTTAGAA	300
	TCAGTGAACA AACGATATGC GCGTGTAGAA AAGATGGCAC GTACGCAAAA AGATAAGAA	360
20	TCAGTAGCAG AATTCAATGT TCTTCAAAAG ATTAACCAG TCCTAGAAGA CGGGAAATCA	420
	GCTCGTACCA TTGAATTAC AGATGAGGAA CAAAAGGTTG TCAAAGGTCT TTTCTTTTG	480
25	ACGACTAAAC CAGTCTTTA TGTAGCTAAT GTGACGAGG ATGTGGTTTC AGAACCTGAC	540
	TCTATCGACT ATGTCAAAAC AATTCGTGAA TTGCAAGCGA CAGAAAATGC TGAAGTAGTC	600
	GTATTCTCTG CGCGTGTGA GGAAGAAATT TCTGAATTGG ATGATGAAGA TAAAAAGAG	660
30	TTTCTTGAAG CCATTGGTTT GACAGAATCA GGTGTAGATA AGTTGACGCG TGCAGCTTAC	720
	CACTTGCTTG GATTGGGAAC TTACTTCACA GCTGGTGAAA AAGAAAGTTC GCGTTGGACT	780
35	TTCAACCGTG GTATGAAGGC TCCTCAAGCA GCTGTATTA TCCACTCAGA CTTTGAAAAA	840
	GGCITTATTC GTGCAGTAAC CATGTCATAT GAAGATCTAG TGAATACGG ATCTGAAAAG	900
	GCCGTAAGAG AAGCTGGACG CTGCGTGAA GAAGGAAAAG AATATATCGT TCAAGATGGC	960
40	GATATCATGG AATTCCGCTT TAATGTCTAA AAATAATAAA ATGGTGTCAA TTAGGTGGA	1020
	AAAAAATTCC AACCTTTTG GCTTTGAAA GGAATAATAA ATGACCAAT TACTTGTAGG	1080
45	TTTGGGAAAT CCAGGGGATA AATATTTGA AACAAACACA ATGTTGTTT TATGTTGATT	1140
	GATCAACTAG CGAAGAAACA GAATGTCACT TTTACACAG ATAAGATATT TCAAGCTGAC	1200
50	CTAGCATCCT TTTTCTAAA TGGAGAAAAA ATTTATCTGT GTTAAACCAA CGACCTTTAT	1260
	GGATTGA	1267

(2) INFORMATION FOR SEQ ID NO:2:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1255 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

60

(ii) MOLECULE TYPE: DNA (genomic)

-41-

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

5

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:2:

10	TGGTCCGTGG TGCTGAGGAC CCTTAGAGTT CGAGTACCAC AAGGTACCGA CTGTTCTGTA	60
	TGCGGAGCTG GCAAGGTTTT AACAGATTTG ATTGAACATG GGCAAGAATT TATCGTTGCC	120
15	CACGGTGCTC GTGGTGAGC TGGAATATT CGTTTCGCGA CACCAAAAAT CCTGCACCGG	180
	AATCTCTGAA AATGAGGAAC CAGGTCAGGA ACGTGAGTTA CAATTGGAAC TAAAAATCTT	240
	GGCAGATGTC GGTTTAGTAG GATTCCCATC TGTAGGGAAG TCAACACTTT TAAGTGTAT	300
20	TACCTCAGCT AAGCCTAAAA TTGGTGCCTA CCACCTTTACC ACTATTGTAC CAAATTTAGG	360
	TATGGTTGCG ACCCAATCCA GGTGAATCCT TTGCAGTAGC CGACTTGCCA GGTTTGATTG	420
25	AAGGGGCTAG TCCAAGGTGT TGTTTGGGA ACTCAGTTC TCCGTCACAT CGAGCGTACA	480
	CGTGTATCC TTCATCATAT TGATATGTCA GCTAGCGAAG GCCGTGATCC ATATGAGGAT	540
	TACCTAGCTA TCAATAAAGA GCTGGAGTCT TACAATCTTC GCCTCATGGA GCGTCCACAG	600
30	ATTATTGTAA CTAATAAGAT GGACATGCCT GAGAGTCAGG AAAATCTTGA AGAATTTAAG	660
	AAAAATTTGG CTGAATAATTA TGATGAATTT GAAGAGTTAC CAGCTATCTT CCCAATTTCT	720
35	GGATTGACCA AGCAAGGTCT GGCAACACTT TTAGATGCTA CAGCTGAATT GTTAGACAAG	780
	ACACCAGAAAT TTTTGCTCTA CGACGAGTCC GATATGGAAG AAGAAGTTTA CTATGGATTT	840
	GACGAAGAAG AAAAAGCCTT TGAATTAGT CGTGATGACG ATGCGACATG GGTACTTTCT	900
40	GGTGAAAAAC TCATGAAACT CTTTAATATG ACCAACTTTG ATCGTGATGA ATCTGTGATG	960
	AAATTTGCCC GTCAGCTTCG TGGTATGGGG GTTGATGAAG CCCTTCGTC GCGTGGAGCT	1020
45	AAAGATGGGG ATTTGGTCCG CATTGGTAAA TTTGAGTTTG AATTGTAGA CTAGGAGACT	1080
	GGTATGGGAG ATAAACCGAT ATCTTTCCGA GATGCGGATG GTAATTTTGT TTCCGCCGCA	1140
	GACGTTTGGA ATGAAAAGAA ATTGGAAGAA CTATTTAATC GTCTCAATCC AAATCGTGCC	1200
50	TTGAGATTGG CACGAACCTAC AAAGGAAAAA CCATCTCAGT AAAGAAGCTA AAAAA	1255

(2) INFORMATION FOR SEQ ID NO:3:

55

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1609 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

60

- (ii) MOLECULE TYPE: DNA (genomic)

WO 98/26072

PCT/US97/22578

- 42 -

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

10	TTACCCATCG CATGACTAAA AATCTCTACT ATCCAATACT AGTTCATATT CTCATCAATA	60
	TCACCTGCCTT CTGGGATGTT TGGTTACTCC TATTTTCAGG AAGTTAGCTT ACTAAAAAAA	120
15	TGTCGGAATT TTCCGGCAIT TTCTTTTTTC ACAAAATAGTC AACGTTTTTC TTTCGGATAC	180
	TGAAGTGGTG TGTAGCCACT TATTTTTTTC AATTGATTTT GAAAATAAGA TTGGCGTGAG	240
20	AAAGGCAGAT AGTGAAGATA GTTAAGAAGA ATAGGATGTT CTTTTTCCT TTTTGGAAAA	300
	CTTCTAAAAT ATGGTATAAT GAAAAGATAA AGAAGTTGGG GGTAGAAGAT GAACATTCAA	360
	CAATTACGCT ATGTTGTGGC TATTGCCAAT AGTGGTACTT TTCGTGAAGC TGCTGAAAAG	420
25	ATGTATGTTA GTCACCGCAG TCTGTCTATT TCTGTCGTG ATTTGGAAAA AGAGTTGGGC	480
	TTTAAGATTT TCGCTCGAG CAGCTCAGGG ACTTCTTGA CCCGTCGTGG GATGGAATTT	540
30	TATGAAAAAG CGCAAGAAIT GGTAAAGGA TTTGATATTT TTCAAAACTA GTATGCCAAT	600
	CCTGAAGAAG AAAAAGATGA ATTTTCGCTT GCTAGCCAGC ACTATGACTT CTTACCACCA	660
	ACTATTACGG CCTTTTCAGA GCGCTATCCT GACTATAAGA ACTTCGGTAT TTTTGAATCA	720
35	ACTACTGTTT AAATATTAGA TGAAGTGGCG CAAGGGCATA GTGAGATTGG GATTATCTAC	780
	CTCAACAATC AAAATAAAAA GGGGATTATG CAACGGGTTG AAAAGTTAGG TCTGGAGGTC	840
40	ATCGAATTGA TTCTTTTCCA TACCCATATT TATCTCTGTG AGGGTCATCC TTTAGCCCAG	900
	AAAGAGGAAT TAGTCATGGA GGATTTAGCG GATTACCAA CGGTTCTGTT CACTCAAGAG	960
	AAAGACGAGT ACCTTTATTA TTCAGAGAAC TTTGTCGATA CCAGCGGTAC TCACAGATGT	1020
45	TTAATGTGAC AGACCGTGCC ACCTTGAAATG GTATTTTGGG GCGGACGGAC GCCTATGCGA	1080
	CAGGTTCTGG ATTTTTAGAT AGTGACAGTG TTAATGGCAT TACAGTTATT CGTCTCAAGG	1140
50	ATAACCTAGA TAACCGCATG GTCTATGTTA AACGTGAAGA AGTGGAGCTT AGTCAAGCTG	1200
	GGACTCTCTT CTGTAAGATC ATGCAAGAAT ATTTTGATCA AAAGAGGAAA TCATGAARAA	1260
	AAGAGCAATA GTGGCAGTCA TTGTACTGCT TTTAATTGGG CTGGATCAGT TGGTCAAATC	1320
55	CTATATGCTC CAGCAGATTC CACTGGGTGA AGTGCGCTCC TGGATTCCCA ATTTCTGTAG	1380
	CTTGACCTAC CTGCAAAATC GAGGTGCAGC CTTTCTATC TTACAAGATC AGCAGCTGTT	1440
60	ATTGCTGTGC ATTACTCTGG TTGTCTGTAT AGGTGCCATT TGGTATTATC ATAAACACAT	1500
	GGAGGACTCA TTCTGGATGG TCTTGGGTTT GACTCTAATA ATCCGCGGTG GTCCTGGAAA	1560





-44-

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:5:

5	GGTCTCTGTTT TGSCCTTGGC GGCTTCAGGT GGTTCAGGAG CTTGGCAGGG AGCTGGTCTC	60
10	ATGTTGGTGT ATACGCTGGG CTTGGCGCTA CCATTCTTGC TTCTAGCTCT CACCTCTAGT	120
15	TATGTTTTGA AACATTTCGG AAAACTTCAT CCCTATCTCG GAATCTCTAA AAAAGTGGGT	180
20	GGTTTTCTCA TTATTGTGAT GGGATTCTTG GTTCTGTTTG GAAATGCTTC AATTTTAAAT	240
25	CAATTATTTG AATAAAATGG AAAGGAATAT CAATATGAAA AATGGCAAA CATGTCTTCT	300
30	TGGAGCAGGT TCGCTCCTTT GTTTCACGGC TTGTCAGGC AAGTCCCTGA CTAGTGAACA	360
35	CCAAACGAAA GATGAAATGA AGACGGAGCA GACAGCTAGT AAAACAAGCG CACTAAAAGG	420
40	GAAAGAGGTG GCTGATTTTG AATTGATGGG AGTAGATGGC AAGACCTACC GTTTATCTGA	480
45	TTACAAGGGC AAGAAAGTCT ATCTCAAATT CTGGGCTTCT TGGTGTTCCT TCTGTCTGGC	540
50	TAGTCTTCCA GATACGGATG AGATTGCTAA AGAAGCTGGT GATGACTATG TGGTCTTGAC	600
55	AGTAGTGTCA CCAGGACATA AGGGAGAGCA ATCTGAAGCG GACTTTAAGA ATTGGTATAA	660
60	GGGATTGGAT TATAAAAATC TCCCAATCCT AGTTGACCCA TCAGGCAAAAC TTTTGGAAAC	720
65	TTATGGTGTG CGTTCTTACC CAACCCAAGC CTTTATAGAC AAAGAAGGCA AGCTGGTCAA	780
70	AACACATCCA GGATTCATGG AAAAAGATGC AATTTTGCAA ACTTTGAAGG AATTATCTTA	840
75	GGAGGCGTCT TATGAATGAT AAGTTAAAAA TCTTCTGTT GCTAGGAGTA TTTTTC	897

## (2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3499 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:6:

55	TTTCTTTTTC CTAGGTGATT TTAATGAGGT TGAATTCAA AATGTATTAG AATCATTGG	60
60	CTTTAAAGGT CGAAAAGGAG ATGTGAAGGT TCAGTATTGT CAACCTTATT CTAATATCCT	120
65	TCAGGAAGGT ATGGTTCGGA AAAATGTGGG ACAATCCATT TTGGAATTAG GTTATCATTA	180

WO 98/26072

PCT/US97/22578

- 45 -

	CCGTTCTAAA	TATGTTGATG	AGCAACATTT	ACCCATGATT	GTAATGAATG	GTTTACTTGG	240
	TGGATTTGGCT	CACCTCTAAGC	TCTTTACAAA	TGTCCTGAA	AATGCTGGAT	TAGCTTATAC	300
5	CATTTCAAGT	GAGCTTGATT	TATTTAGTGG	ATTCCTTGAGG	ATGTATGCTG	GTATCAATCG	360
	AGAAATCGT	AACCAGGCTC	GTAAATGAT	GAATAATCAA	CTGCTTGATT	TAAAAAAGG	420
10	TTATTTTACA	GAGTTTGAGT	TAAATCAGAC	CAAGGAAATG	ATTCGTTGGT	CGTTGTTACT	480
	TTCTCAAGAT	AATCAATCTT	CATTGATTGA	ACGTGCTTAT	CAAAATGCCT	TATTTGAAA	540
	ATCTTCAGCA	GACTTTAAAA	GTTGGATTGC	AAAGCTTGAA	CAAATTGACA	AAGATGCTAT	600
15	TTGTAGAGTA	GCTAATAATG	TGAAACTACA	AGCGATTAC	TTTATGGAAG	GAATAGAATG	660
	ACAAAGGTTG	TTTTTGGAAG	AAAATACTAT	CCAGCTGTAA	AAGAAAAGGT	TTATCGAACT	720
	CGTTTGCCCA	ACGATTGAC	AGTTGCTCTT	TTGCCATAAA	AGGAATTTAA	AGAGGTTTAC	780
20	GGGAGTGTCA	CTGTACAGTT	TGGTTCGGTA	GATACGTTTG	TCACAGAAAT	TGACGGATAT	840
	GTAAACAAT	ATCCTGGAGG	AATTGCTCAT	TTTCTTGAAC	ATAAATTATT	TGAGAGAGAA	900
25	GATTTAGATG	ATTTGATGTC	GGCTTTTACG	AGTCTAGGTG	CAGATAGTAA	TGCTTTTACA	960
	AGCTTTACAA	AAACAACATA	TCTTTTTTCA	GCAACGGATT	ATTTTTTAGA	AAATTTAGAT	1020
	TTACTTGATG	AATTGGTAAC	ATCAGCACAC	TTTACTGGAAG	CTTCATTCT	GACAGAGCAG	1080
30	GATATTATT	AGCAAGAACG	AGAAATGTAC	CAAGATGATC	CAGATTGCTG	TTTATTCTTT	1140
	TCAACTTTAG	CGAATTTGTA	TCCTGGTACA	CCTTTAGCAA	CTGATATAGT	TGGAAGTGAG	1200
35	GAGTCCATTT	CCCAATCAA	TCTAACTAAT	TTGCAAGAAA	ATTTTACAAA	GTTTTACAAA	1260
	CCTGTAAACA	TGCTCTGTT	TTTAGTTGGT	AATTTTGATG	TGGAGCGAGT	ACAGGACTAT	1320
	TTTGAAAGCA	AAGAACTGAA	AGATTGAGAT	TTTCAGGAAG	TAGCAAGAGA	AAAGTTGTTT	1380
40	TTACAGCCTG	TAAAGCCAAC	AGATAGTATG	AGAATGGAAG	TATCTTCTCC	CAAACTAGCG	1440
	ATTGGAGTTA	GAGGTAAAGC	AGAAGTTTCT	GAAGCGGATT	GCTATCGACA	TCATATTTTA	1500
45	TTAAATTTAT	TGTTTGCAAT	GATGTTTGGT	TGGACTTCGG	GATCGTTTTC	AAAAATGTTA	1560
	TGAATCAGGT	AAAATTGATG	CGTCTTATC	TCTGGAAGTT	AAATAACAAG	TCGCTTTCAT	1620
	TTTGTCTATG	TGACAATAGA	TACGAAAGAG	CCAGTTGCTT	TGTCATCA	ATTTAGGAAG	1680
50	GCTATTCGTA	ATTTTACAAA	GGATTTAGAT	ATTACAGAGG	AACATTTAGA	TATTATCAAA	1740
	AGAGAGATGT	TTGGCGAATT	TTTCAGTAGC	ATGAATCTCT	TTGAATTTAT	TGCAACGCAG	1800
55	TATGATGCTT	TTGAAAATGG	TGAGACAATT	TTTGATTGTC	CGAAAATTTT	ACAGGAAAT	1860
	ACTTTAGAGG	ATGTCTCTGA	TGCTGGACAT	CATTTAATAG	ATGATGGTGA	CATAGTTGAT	1920
	TTTACAATAT	TCCCATCGTA	GTAACCTATC	ATAATAGACA	CTAGAAAGAA	GGGATGACAA	1980
60	GTATGAGAAA	AAAAACAATT	GGAGAGGTTT	TACGATTAGC	TAGAATCAAT	CAGGGATTGA	2040

- 46 -

GTTTAGATGA ATTGCAGAAA AAGACAGAAA TCCAGTTAGA TATGTTGGAA GCAATGGAAG 2100  
 5 CAGACGATTT CGATCAACTT CCAAGTCCTT TTTACACGCG TTCTTTCTTG AAAAAATATG 2160  
 CATGGGCTGT TGAGTTAGAT GACCAAAATTG TTTTGGATCG TTATGATTCT GGGAGTATGA 2220  
 TTACTIONTATGA GGAAGTAGAT GTTGATGAAG ATGAGTTGAC AGGTCGTAGA CGTTCAAGTA 2280  
 10 AGAAAAAGAA GAAAAAACA TCATTTTTAC CTTTATTTTA TTTTATCCTT TTTGCTTTAT 2340  
 CGATTTTAAAT TTTTGTGACT TATTATGTTT GGAACATATA TCAAACTCAA CCAGAGGAGC 2400  
 CTTCTCTTTC TAATTACAGT GTGGTTCAAT CAACAAGTTC AACTAGCTCT GTTCCCACT 2460  
 15 CCTCAAGTAG TAGTCTTCT AGTATAGAAT CAGCTATAAG TGTATCAGCG GAAGGAAATC 2520  
 ATGTAGAAAT CGCTTATAAG ACAAGTAAGG AAACAGTTAA ATTGCAATTG GCAGTTTCAG 2580  
 20 ATGTACAAG TTGGGTCAGT GTTTCAGAAA GCGAACTTGA GGGCGGTGTA ACCCTATCGC 2640  
 CAAAGAAGAA AAGTGCAGAA GCAACAGTTG CAACTAAAA TCCTGTAACA ATTACGTTAG 2700  
 GTGTTGTAAG AGGTGTTGAT TTGACAGTAG ATAATCAGAC TGTGATTTA TCGAAATTAA 2760  
 25 CAGCTCAGAC TGGACAATC ACTGTAACCT TTACTIONAAA TTAAGGAAAA ACGAATGAAA 2820  
 AAAGAACAAT TCCCAATCT CTTAACAATA GGTGCAATTC TCTTATACC TATTTTTATC 2880  
 30 TTTATTTTAA CGATAGGAAA TTCGATAGAG AGTCATATAG TTGCACTAT TATCTTTGCT 2940  
 GTTGCCAGTA TTACCGACTA TTAGATGGA TATTTAGCTC GTAAATGGAA TGTGGTCAGT 3000  
 35 AATTTTGTA AATTTGAGA TCCTATGGCG GATAAGTTAC TAGTTATGTC GGCTTTTATT 3060  
 ATGTTGATTG AGTTAGGTAT GGCCTCCGGCT TGGATTGTTG CAGTGATTAT CTGTCGTGAG 3120  
 TTAGCTGTGA CAGGTTTAAG GCTTTTATTG GTTGAAACTG GTGGAACAA TTAGCAGCA 3180  
 40 GCAATGCCTG GAAAAATTAA AACTTTTAGT CAGATGTTTG CTATTATTTT CTGCTATTA 3240  
 CATTGGACTT TGCTTGGTCA AGTTCTACTT TATGTAGCCT TATTTTTAC TATCTACTCT 3300  
 45 GGCTATGACT ATTTCAAGGG TAGTGCCTAT GTATTTAAAG GCACATTGG TTCGAAATGA 3360  
 AATCAATAAT TGATGTAATA AATCTTTCTT TTCGCTATA AGAAAAATCAG AACTACTACG 3420  
 ATGTGAAGGA TATTACGTTT CACGTGAAAC GTGGAGAATG GCTTTCGATT GTAGGGCATA 3480  
 50 ATGGTAGTGG TAAATCAAC 3499

(2) INFORMATION FOR SEQ ID NO:7:

- 55 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 821 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 60 (ii) MOLECULE TYPE: DNA (genomic)

WO 98/26072

PCT/US97/22578

-47-

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

10	ATTTTGAAT AATCAAGCGG AACCAAGAGG TCITCGTCCT TCATCTTGIT AATCATGTAT	60
	TCACTTGGAA TGGCAATATC GTAGGTCGTT CCACCTGCTT TTATCTTAGT GTACATGGCT	120
	TCGTTGGAGT CAAAAGCCTC GTACTGAACT TGAATTCCTG TTTCTTCTGT AAACGTAGTC	180
15	AAGAGTTCAG GATCGATATA GTCTCCCGAG TTATAGATAA CCAATTTTGG ACTATCTCGA	240
	CTATTGAITTT TACTATCTAA ATGAGTCGCA ATCCCCACA AGACAAGGAT AATCGCTGCA	300
20	ATTCCTGCTA AATGAATAG ATTTTITTTCA TGTCTGCTCC TCCTTCTCAC GAGAGATAAA	360
	GTAATAACCT ACAACTAGGA TAATACTAAA GAGAAAGACT AGAGCAGAAA GGGCATTGAT	420
	TTCTAGCGAA ATCCCCCTGC GAGCAGGAGA GTAAATCTCG ACTGATAGGG TTGAAAAGCC	480
25	ATTTCTCTTT ACAAGAAGG TCACGGCAAA GTCATCTAAC GAATAGTGTA AGGCCATGAA	540
	ATAACGACA ATGATAGACG GAGTCAGGTA AGGAAGCATG ATTTCTTAA ACATCTGAAA	600
30	TTGACTAGCT CCCAAGTCAT AGGCCGATG AATCATGTCG CCATTCATTT CCTGAGTCG	660
	GAGGCAAGAC CATCAAGACC ACGATAGGAA TGGAGAAGGC CACATGACTA GATAGAACGG	720
35	TCAAAAAGCC AAGTGAAAAC TTGAGTTGGG TAAAGAGAAT CAAGAAGTAG CACCAATCAT	780
	AACGTCAGGC GCAACCATGA GGATATTATT GAGTGATAGA A	821

(2) INFORMATION FOR SEQ ID NO:8:

40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1309 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

50 (iv) ANTI-SENSE: NO

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

60	GGCCGGTGCC ACAGTCCAAG CTATCGGTAT CGTGATTGAG AAATCCTTCC AAGATGGTCG	60
	TGATTTGCTT GAAAAAGCAG GCTACCTGT CCTATCACTT GCTCGTTGG ATCGTTTGA	120
60	AAATGGTCAG GTCGTATTTA AGGAGGCAGA TCTCTAATGC AAATCAAGA AAAACACTCG	180

- 48 -

	CAAGCAGCCG TTCTGGGCTT GCAGCACTTA CTAGCCATGT ACTCAGGATC TATCCTGGTT	240
	CCCATCATGA TTGCGACAGC CCTTGGCTAT TCAGCTGAGC AGTTGACCTA CCTGATTCT	300
5	ACAGATATCT TCATGTGTGG GGTGCAACC TTCTCCAAC TCCAACCTAA CAAATACTTT	360
	GGGATTGGAC TCCAGTCGT TCTTGGAGTT GCATTCCAGT CGGTGCTCC CTTGATTATG	420
10	ATTGGGCAA GCCATGGTAG TGGCGCTATG TTTGGTGCC TTATCGCATC TGGGATTTAC	480
	GTGGTTCTTG TTTAGGCAT CTCTCAAAA GTAGCCAATC TCTTCCCATC TATCGTAACA	540
	GGATCTGTTA TTACCACGAT TGGTTTAACC TTGATCCCTG TCGCTATTGG AAATATGGGA	600
15	AATAACGTTT CAGAGCCAAC TGGTCAAAGT CTCTTGCTTG CAGCTATTAC TGTCTTGATT	660
	ATCCTCTTGA TCAACATCTT TACCAAAGGA TTTATCAAGT CTATCTCTAT TTTGATTGGT	720
20	CTGGTTGTGG GAACTGCCAT TGCTGCTACT ATGGGCTTGG TGGACTTCTC TCCTGTTGGC	780
	GTAGTCCACT TGTCCATGTC CCAACTCCAC TCTACTTTGG GATGCCAACC TTTGAAATCT	840
	CATCTATTGT CATGATGTGT ATCATCGCAA CGGTGCTCTAT GGTGAGTCA ACTGGTGTTT	900
25	ATCTAGCCTT GTCTGATATC ACAAAGATC CAATCGACAG CACGGGCTT CGCAACGGTT	960
	ACCGCGCAGA AGGTTTGGCG GTACTTCTCG GAGGAATCTT TAACACCTTC CCTTACACCG	1020
30	GATTTTCACA AAACGTTGGT TTGGTTAAAT TGTCAGGCAT CAAAAACGC CTGCCAATCT	1080
	ACTACGCAGG TGGTTTCCTG GTTCTCCTTG GACTGCTTCC TAAGTTTGGT GCCCTTGCCC	1140
	AAATCATTCG AAGCTCCGTC CTCGGCGGTG CCATGCTGGT GATGTTTGGT TTTGTATCTA	1200
35	TTCAAGGGAT GCAATCCTC GCCCGAGTTG ACTTTGTAAC AATGAACACA ACTTCCTTAT	1260
	CGCAGTGTTT CAATCGCTGC AGGTGTCGGT CTCAACAACA AGTAATCTC	1309

40 (2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1031 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

50 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

60	TTAAAGTTC AGTTTATCTA GGTCTTCAT TTGCCTTTAT CACAGCTATG TCATGGCTA	60
	TGAAGAAAT GGGGGTGAT GTATCTGCTG CCCAACAGG GGTATCTTG ACTGGTTTGG	120

WO 98/26072

PCT/US97/22578

- 49 -

	TCTATGCTCT TGTGCTACC AGCATCCGAT TTGTAGGAAC AAAATGGATT GATAAACTCT	180
5	TGCCACCAAT CATTATCGGT CCTATGATCA TCGTTATCGG TCTTGGACTT GCAGGTTTCTAG	240
	CTGTTACCAA TGCAGTGCTT GTAGCAGACG GAAATGGAA AAATGCTCTG GTAGCCGTTG	300
	TTACTTTCTT AATTGCTGCC TTTATCAATA CAAAAGGAAA AGGCTTCCTA CGAATCATTC	360
10	CATTCTCTT TGCCATTATC GGTGGTTACC TTTTCGCACT AACTCTTGGC TTGTTGACT	420
	TTACACCACT TCTTAAAGCC AACTGGTTCG AAATTCCTGG TTTCTACTTG CCATTTAGCA	480
	CAGTGGTGC CTTTAAAGAG TACAATCTT ACTTTGGTCC AGAAGCCATC GCTATCTTGC	540
15	CAATCGCTAT CGTAACAATT TCTGAACATA TCGGAGACCA TACTGTTTTG GGTCAAATCT	600
	GTGGCCGTCA ATTCTTAAAA GAACCAAGGT TCCATCGTAC TCTTCTTGGT GACGGTATCG	660
20	CAACTCTGT TTCTGCCTTC CTGGTGGAC CAGCCAATAC AACTTACGGA GAAAATACAG	720
	GGGTTATCGG TATGACTCGT ATCGCTTCTG TCTCAGTTAT CGGTAACGCT GCCTTCATCG	780
	CGATTGCCCT CAGCTTCTTT GGTAAATTCA CTGCCCTTAT TTCAACTATT CCAAACGCTG	840
25	TACTTGGTGG TATGCTCAATC CTTCTCTATG GGGTTATCGC CAGCAATGCT TTGAAAGTCT	900
	TGATTAAAGA ACGTGTGAT TTCGCTCAAA TCGGAACCT CATCATCGCA AGTGCTATGT	960
30	TGGTCTTTGG ACTTGGAGGA GCTATCCTTA AACTTGGTCC AGTACACTTT CAGGTACTGC	1020
	CCTTTCAGCC A	1031

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 568 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

	ACAGTTTAAAT CATTGCCTTG GCTACAACCC TCATTGCGAT TATTATTCTT GCTATGGCAG	60
55	CCTATGGTAT TGTTCGATT TTTCTAAAT TGGGAGCAAT CATGTCGAGA CTACTCGTCA	120
	TTACCTACAT TTTCCACCA ATTTTGTTAG CAATTCCCTA TTCAATTGCC ATTGCTAAAG	180
	TTGGGTTAAC AAATAGTTTA TTGGCTTGA TGATGGTTTA TCTATCTTTT AGTGTTCAT	240
60	ATGCAGTTTG GCTCTTAGTT GGATTTTTC CAAACAGTCC AATTGGAATT GAAGAAGCGG	300

WO 98/26072

PCT/US97/22578

-50-

	CTAGAATTGA TGGTGCAAAT AAATTGTGTA CGTTTTATAA AGTTGTGCTA CGGATTGTAG	360
5	CACCAGGTAT TGTAGCAACA GCTATTATTA CATTATCAA TGCTTGAAT GAATTCTGT	420
	ATGCCITGAT TTGATTAAC AATACAGGAA AGATGACAGT AGCAGTAGCC CTTCSTTCAC	480
	TTAATGGTTC AGAATACTA GACTGGGGAG ATATGATGGC AGCGTCTGTT ATTGTAGTTC	540
10	TTCCATCAAT TATTCTTCT CTATCACC	568

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 468 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:	
	ACCAAAGACT TAGCTTCTTC AAAAAGCGGA TCACCACCAG CATCTCCATC CGAAAATTCT	60
	CCTTCATTT CAGAAACCTC ACCTGGATCA AAACCTCAT CGTAGCTGC ATCTGCCTGA	120
35	GTCTTGATGA AGTTCACAAT GCGCTCAACA TCGTCATCCG AGATAAAGGA GCCTTGGAGA	180
	CGAACTGGAT GATTTTCATC AATCGGTTTA AAGAGCATGT CTCCTCGACC AAGAAGTTT	240
40	TCTGCTCCAT TTTCATCAA AATCGTACGG GAGTCTGTC CTGATGAAC CGCAAATGCT	300
	ACACGAGATG GAACATTGGC CTTAATCAA CCAGAGATGA CATCAACAGA TGGACGCTGA	360
	GTTGCAAGAA TCATGTGGAT ACCTGCAGCA CGCGCTTCT GCCCAAGACG GATGATAGCA	420
45	TCTTCCACTT CCTTCTGTCG CACCATCATG AGTCAAGCCA ACTCATCC	468

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 466 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

60



WO 98/26072

PCT/US97/22578

-51-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

5	AAGCTGACAA TCTTTCTGC AGTTGGAGCA TCCGAGAAGG ATACACCACT AAGGATGCGA	60
	CCTGCCTTGC TATCAACAAAT AATGTCTTGA ACCTTGTAGT CATCTCCATA GACCAAGAAG	120
	CATTGCTTGG TACAATCTTC ACGATAAACA CTAAAATAAG TCGAACGAGT CAAATCATTG	180
10	CGGAACATAT TTTTAAAGAG ATAGTTATCT GCATCAATAA CATAGCTGTT GGCCAATTCT	240
	TCTTTTACAA GATAGAGAGA GTAAAAGTTA TTGTAGTCAG CGTATTATC ATTGAAAACG	300
15	AGACGAACAC CGTATTCTC TTCAAGTAA TCGAATTGTT CTTTAAAGATA ACCAACAATG	360
	ATGATGATGT CATTGATTCC TTTTCTTTG AGAACTCAA TTTGGTACTC AATCAAAGGT	420
	TTTTGATTAA CCTGAACCAA GGCCTTAGGG GTATTTCAG TCATAG	466

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

25	(A) LENGTH: 1040 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: single
	(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

40	CACATAATCT GTATATTGAC TATAAGTTTT AAAAAACAAT TTTTAAGCTC TTCCTTGCT	60
	TCTCTAACCA AGCGTGTAT AATGAATACT GCTCAAGCGA CCTTCAATCG TGAAGCACAC	120
	ACGACCTTCA ATCGTGAATA AACGAATAGA TGGGAGACTT ACCATGAGTG ATAACCTTAA	180
45	AACACGTGTT GTCTGGGGA TGAGTGGTGG TGTGATTGCG TCGGTGACGG CTCCTTTGCT	240
	CAAGGAGCAG GGCACGATG TGATCGGTAT CTTTATGAAG AACTGGGAGT ACACAGATGA	300
50	AACCGCGCTC TGTACGGCGA CCGAAGATTA CAAGGATGTG GTTGCCTGCG CAGATCAGAT	360
	TGGCAITCCC TACTACTCTG TCAATTTTGA AAAAGAGTAC TGGGACCGCG TTTTGTAGTA	420
	TTTCCTAGCT GAATACCGTG CAGGGCGCAC GCCAAATCCG GACGTTATGT GCAACAAGGA	480
55	AATCAAGTTC AAGGCCCTTT TGGACTATGC CATGACCTTG GGGGCAGACT ATGTAGCGAC	540
	TGGGCACTAT GCTCGAGTGG CGCGTGATGA GGATGGCACT GTTCACATCG TTCGTGGCGT	600
60	GGACAATGGC AAGGATCAGA CCTATTTCCT CAGCCAACCT TCGCAAGAAC AACTTCAAAA	660



WO 98/26072

PCT/US97/22578

-53-

	TTTTTAAAAA AGCACGAGGT TTCTAAGGGA TTGCTGGCCA AGATTAAGTT TCGAGGTGGA	900
	GCTATTCTGG TCAATAATCA ACCGCAAAAT GCACGSTATC TATTGGACGT TGGAGACTAC	960
5	GTACCAATTC ACATTCCCGC TGAGAAAGGC TTGAAACCT TGGAGGCTAT TGAGCTTCCA	1020
	TTAGATATTG TCTATGAGGA TGACCACCTT CTAGTCTTGA ATAAACCCCTA TGGAGTGGCT	1080
10	TCTATTCTTA GTGTTAATCA CTCTAATACC ATTGCGAATT TTATCAAGGG TTACTATGTC	1140
	AAGCAAAATT ATGAAAATCA GCAGGTTTAC ATTGTTACCA GACTAGATAG GGACACTTCT	1200
	GGCTTGATGC TCTTTGCCAA GCACGGTTAT GCCCATGCAC GATTAGACAA GCAGTTGCAG	1260
15	AAGAAATCTA TCGAGAAACG CTACTTTGCT TTGTTAAGG GAGATGGACA TTGGAGCCA	1320
	GAAGGGGAAA TTATTGCTCC GATTGCGCGT GATGAAGATT CCATTATTAC CAGACGAGTG	1380
20	GCTAAAGGCG GAAAGTATGC CCATACTTCA TACAAGATTG TAGCTTCTTA TGGAAATATT	1440
	CACCTGGTCT ATATTCACTT GCACAGTGGT CGAACCCATC AATCCGAGT CCATTTTCT	1500
	CATATCGGTT TTCTTTTGGT GGGAGATGAT TTGTATGGTG GTAGTCTGGA AGATGGTATT	1560
25	CAACGTCAGG CTCTGCATTC CCATTACCTA TCCTTTATTC ATCCATTTT AGAGCAAGAC	1620
	TTGCAGTTAG AAGTCCCTT GCCGGATGAT TTCAGTAACC TTATTACCCA GTTATCAACT	1680
	AATACTCTAT AAAAAGTCTC TCAGAGTATA ATTATTATCT TAAAGGAGAA AACTCATGGA	1740
30	AGTTTTTGAA AGTCTCAAAG CCAACCTTGT TGTTAAAAAT GCTCGTATCG TTCTCCCTGA	1800
	AGGGGAAGAG CCTCGTATTC TTCAAGCAAC AAAACGCTTA GTAAAAAGAA CAGAAATGAT	1860
35	TCCTGTTTTG CTTGAAATC CTGAAAAAAT TAAAAATTAT CTTGAAATTG AAGGAATCAT	1920
	GGATGGTTAT GAGGTATCG ACCCTCAACA TTATCCTCAA TTTGAAGAAA TGGTTTCTGC	1980
40	CTTGGTGGAG CGTCGCAAGG GCAAAATGAC TGAAGAAGAT GTACGCAAGG TTTTGGTTGA	2040
	AGATGTCAAC TACTTTGGTG TGATGTTGGT TTAAGTGGG TTGTTGATG GAATGCTGTC	2100
	AGGAGCGATT CACTCAACAG CTTCAACAGT TCGCCAGCT CTACAAATCA TCAAACTCG	2160
45	TCCAATGTA ACTCGTACTT CAGGAGCCTT CCTCATGGTT CGTGGTACGG AACGTTACCT	2220
	ATTTGGAGAC TGTGCCATTA ATATCAATCC AGATGCAGAA GCCTTGGCTG AAATTGCCAT	2280
50	CAACTCAGCA ATCAGAGCTA AGATGTTTGG CATCGAACCT AAAATTGCCA TGTGTAGCTA	2340
	TTCTACTAAA GGTTCAGGCT TTGGTGAAG CGTTGATAAG GTCTGTGAAG CAACATAAAT	2400
	TGCTCAGCAG TTGCGTCTG ACCTTGAAAT CGATGGTGAG TTGCAATTGG ATGGGCGCTT	2460
55	TGTTCCCGAA ACTGCAGCTC TGAAGCTCC GGGAAAGTACA GTAGCTGGTC AAGCAATGAT	2520
	CTTCATCTTC CCAGGTATCG AGGCAGGAAA TATCGGTTAC AAGATGGCTG AACGCTGGG	2580
60	TGGCTTTGCG GCTGTAGGAC CTGTTTTCGA AGGTTTAAAC AAGCCAGTTA ATGATCTTTC	2640
	TGCTGGATGT AATGCAGATG ATGTTTACAA GTTGACCCCT ATCACAGCAG CTCACAGCAT	2700

WO 98/26072

PCT/US97/22578

-54-

TCATCAATAG TGAAGACTAT AAAGTGATAT ACTATGCTAT ACTGTAGTTA TGAAACTATG 2760  
 TACGAAAAGC ACTGCCATTA ATTCCTGAGA ACTAAATTAC TGATTGGTGT CAAAAAGSAA 2820  
 5 AACTTCCAAG CGATGATATC CTGTCTATAC ACGACCTATA GAAATCTGTA ATATACATGT 2880  
 CCGTAAAGC ATAAATTCCC TTTTGTATT TAAATGAGTA TGAAGAGAGA ATTTCCGGC 2940  
 10 TCTTTGCAA CTGTAGTGGG TTGAAAAA GCTAAGCTCG AGAAGGACA AATTTTGTCC 3000  
 TTTCTTTT TATATTGAGA GCGATAAAA TCCGTTTTT GAAATTTTCA AAGTTTCGAC 3060  
 TCTAGAGGAT C 3071

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 720 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TTTCCATGGT ATGGTAAAGG TTTTCTTTT TTTTAAAGG AAAACGAGAA GAGGAGGTTT 60  
 TTATGAAAGC AAGCATTGCC TTGCAAGTTT TACCCCTAGC ACAGGGGATT GATCGGATAG 120  
 CTGTATTGA TCAGGTCATT GCTTATCTGC AACTCAAGA AGTGACGATG GTAGTGACAC 180  
 40 CATTTGAAAC GGTCTTGAA GGGGAGTTT ATGAGCTTAT GCGCATTCTA AAGAGAGCGC 240  
 TGGAASTGGC AGGCGAGGAG GCAGACAATG TCTTGCCAA TGTCAAAATA AATGTAGGAG 300  
 45 AGATTTTAA GATTGATGAG AAACITGAAA AGTATACTGA GACGACACAT TAGTCTATTG 360  
 GGCTTCTCG GAGTATTGTC AATCTGGCAG TTAGCAGGTT TTCTTAACT TCTCCCAAG 420  
 TTTATCTGC CGACACCTCT TGAATTTCTC CAGCCCTTGG TTCTGAGAC AGAATTTTCT 480  
 50 TGGCACCATA GCTGGGCGAC CTTGAGAGTG GCTTACTGG GGCTGATTTT GGGAGTTTTC 540  
 ATTGCTCTGC TTATGGCTGT GCTCATGGAT AGTTTGACTT GGCTCAATGA CCGATTTCAC 600  
 55 CCTATGATGG TGGTCATTCA GACATTCG ACATTGCCA TAGCTCCTAT CCGGTCTTG 660  
 TGGCTGGTT ATGGGATTTT TGCCCAAGAT TGTCTTGATT ATCTTACAA CAACCTTTCC 720

(2) INFORMATION FOR SEQ ID NO:16:

-55-

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 852 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:
- |                                  |     |
|----------------------------------|-----|
| GCCGTCATAA TCATGCGCCG AATCCGTC   | 60  |
| CCATTAAATC TGGGTCGTGA AAGACAATGA |     |
| CTCCATGACG TTGGTGATGA CGCTGAATCC | 120 |
| GCTCTATGTC CTGGTCATTG ATGGCAGAAC |     |
| CTCGAGTCTC ATAGGTCCTC ACATCGAAAT | 180 |
| AACGCTTGAG ATTGACCGTA TCATCAGGAC |     |
| CTTCAACCAC GATAACTTGG GAAATTCTCT | 240 |
| CTTTCAATAC TTGCTGTCCA ATCCCAAAAA |     |
| TGCGTTCTGC ATTTGCAGTC GTTGCTACCG | 300 |
| CCAGCTCTTC TGTGTCATA CCACGCAAGT  |     |
| CAGCGATAAA GTCGACCACA TAGCGAGTAT | 360 |
| AGGCTGTTTT ATTTTCACGA CCACGCTTGG |     |
| GTACAGGTGC TAAGTAAGGC GCATCTGTTT | 420 |
| CTACCAACAT CTTGTCCAAA GGTAACTCTT |     |
| TAGCTGCTTC TTGAGGTCA GTTGCTTCT   | 480 |
| TGAAGTCAC CACTCCTGAG AAGGAAATGG  |     |
| TCATACCAAG ATCCCGGTAC CGAGCCCACT | 540 |
| CAAGGTCGCC TGAAAATGAA TGCATGATAC |     |
| CACCACGAGG ACCAACGCCC TCACCTCTGA | 600 |
| TAACTCTATA GGTATCTTCC AGCGCATCAC |     |
| GGGTATGGAC AAAAAAGGC AAATCCAAGT  | 660 |
| CCTTAGATAG CTGAATCTGA CGGCGAAAAA |     |
| CTGTCTCTTG CACCTCTTGG GCGGTGCAT  | 720 |
| CCAATGGTAG TCTAAGCCAA TTTCACCTAA |     |
| AGCCACAACC TTGGAATGTT TTAAGTTATC | 780 |
| CAACAAGTAA GCCTCAACTT CCTCTGTATA |     |
| AGTACCAGCT TCTGTAGGAT GCCAACCAAT | 840 |
| AGTCGCATAG AGCTGCTCAT ACTCATCTAC |     |
| CAAACTCCAA GG                    | 852 |
- (2) INFORMATION FOR SEQ ID NO:17:
- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 66 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

## (xi) SEQUENCE DESCRIPTION: SEQ ID No:17:

5 GGGGATCCTC TAGAGTCGAT ATCTACGGTC TCAACCGTAC AGGACTGTTG AACGATGTAC 60  
TGCAAGTTCT TTCAAATAGA ACCAAGAATA TTTCACGGT CAATGCCCAA CCAACCAAGG 120  
10 ATATGAAGTT TGCTAATATC CATGTCTCCT TCGGTATTGC CAACCTCTCT ACACTGACCA 180  
CGGTGTGCGA TAAATTAAG AGTGTGCCAG AAGTTTACTC TGTCAAACGG ACCAACGGCT 240  
15 AGTTGTCCTA GCTCTTACTA GAAAGGCTAT TATGAAAATC ATTATCCAAC GGGTTAAAAA 300  
AGCCCAAGTG AGTATAGAAG GCCAGATTCA GGGAAAAATC AATCAGGGAC TTTTATTGCT 360  
GGTTGGTGTT GGACCAGAGG ACCAAGAGGA AGATTTGAC TATGCTGTGA GAAAACTGGT 420  
20 CAATATGCGG ATTTTTCAG ACGCAGAAGG CAAGATGAAC CTGTCTGTCA AAGATATTGA 480  
AGGAGAAATC CTCCTATTT CTCAGTTTAC CCTCTTTGCG GATACTAAGA AAGGCAATCG 540  
TCCAGCCTTT ACAGGGGCGAG CTAAACCTGA TATGGCATCA GACTTCTATG ATGCTTTCAA 600  
25 TCAAAAATTA GCGCAAGAAG TGCCCGTTCA GACAGSTATC TTTGAGCAG ATATGCAGGT 660  
TGAGCTGGTT AATAACGGAC CTGTACCAT TATCCTAGAT ACTAAAAAGA GATAAGAAAG 720  
30 ACCAAGCCCA GTCGGCTTGG TCTTTCTCAT CGATCATAAA AATACTCCAA AAAGAAATCG 780  
GTTCTTGATA TGCTTGGGGG ACTCTTTTCA GGCTTTGGCA GATGCGATAG GAAGGGATGA 840  
35 GATGTCTTAG GGTGAGGAGA GTTCCCTG 868

## (2) INFORMATION FOR SEQ ID No:18:

## (i) SEQUENCE CHARACTERISTICS:

40 (A) LENGTH: 1399 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

45 (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

50

## (xi) SEQUENCE DESCRIPTION: SEQ ID No:18:

55 CGGTCTCTGT CCGATTGACT CACACCTTAA GCGCTTTGAA GCTATGGGTG CCACTGCTAG 60  
CTACAGGGGA GATAACATGA AGTTATCTGC TAAAGATACA GGACTTCATG GTGCAAGTAT 120  
60 TTACATGGAT ACGGTTAGTG TGGGAGCAAC GATTAAATACG ATGATTGCTG CGGTTAAAGC 180  
AAATGGTCGT ACTATTATTG AAAATGCAGC CCGTGAACCT GAGATTATTG ATGTAGCTAC 240

WO 98/26072

PCT/US97/22578

- 57 -

	TCTCTTGAAT AATATGGGTG CCCATATCCG TGGGCGAGGA ACTAATATCA TCATTATTGA	300
5	TGGTGTGTGAA AGATTACATG GGACACGTCA TCAGGTGATT CCAGACCGCA TTGAAGCTGG	360
	AACATATATA TCTTTAGCTG CTGCAGTTGG TAAAGGAATT CGTATAAATA ATGTTCTTTA	420
	CGAACACCTG GAAGGGTTTG TTGCTAAGTT GGAAGAAATG GGAGTGAGAA TGACTGTATC	480
10	TGAAGACAGC ATTTTGTGCG AGGAACAGTC TAATTGAAA GCAATCAATA TTAAGACAGC	540
	TCCTTACCCA GGCTTTGCAA CTGATTGCA ACAACCGCTT ACCCCTCTTT TACTAAGAGC	600
	GAATGTCGT GGTACAATTG TCGAGTCGAT ACGATTACG AAAAACTGT AAATCATGTT	660
15	TTTGAAC TAG CAAAGATGGA TCGGATATT TCGCAACAA ATGGTCATAT TTTGTACAGC	720
	GGTGGACGTG ATTTACGTGG GGCCAGTGT AAAAGCAGCG ACTTAAGAGC TGGGGCTGCA	780
20	CTAGTCATTG CTGGGCTTAT GGCTGAAGGC AAAACTGAAA TTACCAATAT CGAGTTTATC	840
	TTACGTGTTT ATTCTGATAT TATCGAAAAA TTACGTAATT TAGGAGCGGA TATTAGACTT	900
	GTTGAGGATT AAACCGTAGA GGTGTTTATG AATATTGGA CCAATTAGC AATGTTTTCT	960
25	TTTTTTGAAA CGGATCGCTT GTATTTGCGT CCTTCTTTT TTAGTGATAG TCAGGACTTC	1020
	CGCGAGATAG CTTCAAATCC AGAAAATCTT CAATTATATT TCCCAACGCA GGCAAGCTG	1080
30	GAAGAAATC AATATGCACT GGCCAATTAC TTTATGAAT CCCCTTGGG AGTGTGGGCA	1140
	ATTTGTGACC AGAAAAATCA ACAATGATT GGTCTATTA AATTTGAGAA GTTAGATGAA	1200
35	ATCAAAAAAG AAGCTGAGCT TGGCTATTTT TTGAGAAAA ATGCTTGCTC GCAAGGATT	1260
	ATGACAGAGG TTGTTAGAAA AATTTGTCAG CTTTCTTTT AGGAATTTGG CTTAAAACAA	1320
	TTATCTATCA TTACCCACCT TGAAAAATGA GCTAGCCAAA GAGTTGCTCT TAAGCTGGA	1380
40	TTTAGTTTGT TCGCTCAGT	1399

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1779 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

60	AGATTGCTCT TGAACAGCAT GAAATACCAA TTGTTGCTGT GATTGTCAAA GATGGGAAAA	60
----	---	----





-59-

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3725 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID No:20:

GCGGATCCTC	TAGAGTCGAA	AGATTACGAA	GGTAAGAAC	CTCTTTATTA	CTGGTCACAT	60
CATGCTACAA	CAAGCTGCAA	CAGTATCTCT	TATGGTTCTA	TTCTTAGTAC	CACAATTGCG	120
CAATGCTTAC	GGTACAGCAG	CGATTGGTAT	CATCTGTGGA	CTTTACTGGG	CAGTTAGTTC	180
AAATATGACT	GTTAGGCCAA	CTCAACGCTT	GACTGTGGT	GGCGGATTGG	CGATTGTGCA	240
CCAACAGCAA	TTTGCAATCT	GGTTGTAGA	TAAAGTAGCA	GGACGCTTTG	GTAAGAAAGA	300
AGAAAGTTTA	GACAATCTTA	AATTACCTAA	GTTCCTCTCA	ATCTTCCACG	ATACAGTTGT	360
TGCATCTGCT	ACCTTGATGC	TCGTATTCTT	CGGGGCCATT	CTTTTAATCT	TGGGTCCAGA	420
CATTATGTCT	AATAAAGAAG	TCATCACTTC	AGGAACCTTA	TTCAATCCTG	CTAAACAAGA	480
TTTCTTTATG	TACATTATCC	AAACAGCCTT	TACCTTCTCA	GTTTACTTGT	TCGTTTGTAT	540
GCAAGGTGTC	CGAATGTTGG	TATCTGAGTT	AACAAACGCT	TTCCAAGGTA	TTTCAAACAA	600
ATTGTTGCCA	GGTTCATTCC	CAGCGGTTGA	CGTTGCAGCT	TCTTATGGAT	TTGGTTCTCC	660
AAATGCTGTC	TTGTCAGGAT	TTACCTTTGG	TTTGATTGGT	CAATTGATTA	CAATTGTCTT	720
GCTCATCGTC	TTTAAAAATC	CGATTCTTAT	TATTACAGGA	TTTGACACAG	TGTTCTTTGA	780
CAATGCAGCC	ATTGCGGTCT	ACGCTGATAA	ACGCGGCGGA	TGGAAAGCGG	CTGTTATCCT	840
TTCCCTTTATA	TCAGTGTGCC	TTCAAGTTGC	TCTAGGAGCT	CTTTGTGTGG	CCCTTCTCGA	900
TTTGCACTCT	TATGGTGGCT	ACCATGGAAA	TATCGACTTT	GAATTCCCAT	GGCTTGGATT	960
TGGATATATC	TTCAAATACC	TTGGTATTGT	TGGTTATGTA	CTTGTGTGTC	TCTTCTTGCT	1020
TGTTATTCCT	CAACTTCAAT	TTGCCAAAGC	AAAAGATAAA	GAGAAATATT	ACAACGGTGA	1080
AGTTCAAGAA	GAAGCTTAGT	ATCTAGAAAA	GGAGAAATAA	AATGGTTAAA	GATTATAGCAG	1140
CGTGGGAAA	TGGAATGGGT	TCATCAATGG	TTATCAAGAT	GAAGGTTGAA	AATGCTCTCC	1200
GTAAGCTTAA	TCAAACAGAT	TTTACAGTCA	ATTCATGCAG	TGTCGGTGAA	GCTAAAGGTT	1260
TAGCAGTAGG	ATATGACATC	GTAATCGCTT	CTCTTCATT	GATTCAAGAA	TGGGAAGGGC	1320

WO 98/26072

PCT/US97/22578

- 60 -

	GAACATAATGG GAAGTTAATT GGAATTGATA ACTTGATGGA TGATAAAGAA ATCACCAGAAA	1380
	AACCTAGTCA AGCACTACAG TAAAAGGTTG GAGGSGCTG GACAGAACT GAGAGTTATC	1440
5	GTTTCTGTCC TTCTCCCTCT TTAATAAAG GAGGCAGATA TGAATTTAAA ACAAGCTTTA	1500
	ATTGACAATG ACTCGATCCG ACTAGGTTTA GAGGCTAACA ATTGGAAAAG AGCAGTCAAG	1560
	GTAGCAGTAG ATCCCTTAAT TGAAGTGGG GCAATTTTGC CAGACTATTA CGATGCTATC	1620
10	ATTGAATCGA CTGAAGATA TGGGCTTAC TATATCTTGA TGCCAGGTAT GGTATGCCC	1680
	CACGCTAGAC CTGAAGCTGG TGTGCAAGT GATGCCTTTT CATTGATTAC CTTACAAAAT	1740
15	CCTGTGTAT TTTGAGATGG GAAAGAGGTA TCTGTTTTGT TGGCACTAGC AGCAACAAGC	1800
	TCAAAATTC ACACAAGTGT AGCCATTCCA CAATATTGTT CCCTGTTTGA ATTAGAAGAT	1860
20	TCTATTGCAC GTTTACAGGC TTGCCAGACT AAGAAGATG TCTTGCTAT GATTGAAGAA	1920
	TCTAAGGATA GCCCTTATCT CGAAGGATTG GATTGGGAAA GTTAGAAAAG GGAATAAAGA	1980
	AATGACAAAA AGAATACCTA ATTTACAAGT TGCATTAGAC CATTGAGACT TGCAAGGAGC	2040
25	GATTAAAGCA GCTGTTTCTG TTGGTCAGGA AGTAGATATT ATCGAAGCTG GAACGTGTTG	2100
	CTTGCTTCAA GTTGGAAAGT AACTGGCTGA AGTCTTGCCT AGCCTTTTCC CAGATAAGAT	2160
30	TATTTGGGCA GACACAAATG GTGCTGATGC TGTGGAACA GTTGCTAAAA ATAATGCGGT	2220
	TCGTGGAGCA GACTGGATGA CTTGTATCTG TTGTGCAACC ATCCCTACTA TGGGAAGCAGC	2280
	TCTAAAGGCT ATCAAGACTG AACGAGGAGA ACGAGGCGAA ATCCAGATCG AGCTTTATGG	2340
35	CGATTGGACT TTTGAACAAG CTCAGCTTTG GCTAGATGCA GGTATTTTAC AAGCTATTTA	2400
	TCACCAATCT CGTGATGCTC TTCTTGTCTG TGAAACTTGG GGTGAAAAAG ACCTTAATAA	2460
40	GGTTAAAAAA CTCATTGACA TGGGCTTCCG TGTATCTGTA ACAGGTGGTC TAGATGTAGA	2520
	TACTCTCAAA CTCCTTGAAG GTGTTGATGT CTTTACCTTT ATCGCAGGTC GTGGAATTAC	2580
	AGAGGCTGCG GATCCAGCAG GAGCAGCGCG TGCCCTCAAG GATGAAATCA AACGAATTG	2640
45	GGGGTAAATC ATGGTACGTC CAATTGGAAT TTATGAAAAG GCAACCCCAA CACACTTTAC	2700
	TTGGCTAGAA CGTTTAAATT TTGCCAAGGA GTTAGGCTTT GATTTTGTG AGATGTCTAT	2760
50	TGACGAACGT GACGACGCTT TAGCAAGACT TGACTGGAGT AAGGAAGAAC GCTTGGAAAT	2820
	TGTCAAAGCA ATCTATGAAA CTGGTGTTCC TATTCCTCTT ATCTGTTTTT CAGGCCATCG	2880
	TCGCTACCCA TTGGGTTCAA AAGATCCAGT TCTAGAGGAA AAATCTCTAG AACTCATGAA	2940
55	AAAATGTATC GAATTAGCTC AAGACTTGGG AGTTCGTACG ATTCAATTAG CTGGTTACGA	3000
	TGTTTACTAT GAGGAAAAAT CACCCAGAC ACGCCAACGT TTTATCAAAA ATTTGAGAAA	3060
	AGCCTGTGAC TGGGCTGAAG AAGCTCAGGT GGTACTTGCT ATTGAAATTA TGGATGATCC	3120
60	TTTCATCAAT AGCATCGAAA AATATTGGC TATAGAAAA GAGATTGACT CTCCTTCTCT	3180

-61-

	CITTTGATAT	CCAGATATTG	GTAATGTGTC	TGCATGGCAT	AATGATATCT	ATAGTGAGTT	3240
5	TTATCTTGGT	CATCATGCCA	TGCGAGCTCT	CCATCTCAAG	GATACTTATG	CAGTACACAGA	3300
	AAGTTCAAAG	GGCCAGTTCC	GAGATGTACC	TTTCGGGCAA	GGTGTGTCA	AATGGGAAGA	3360
	AGCTTTCGAT	ATTITAAAGG	AAACCAATTA	TAATGACCTT	TTCTTAATCG	AAATGTGGTC	3420
10	TGAAATTTGT	GAACAGTAG	AAGAAACACG	CGCAGCCGTT	CAAGAGCGCG	AAGCTTTTCT	3480
	CTATCCACTC	ATTAAGAAAG	CAGGTTTGAT	GTAAGATGAA	TCAAGTAATC	AATGCTATGC	3540
15	GTAACAGAGT	CTGTGATGCC	AATCAATCAT	TGCCAAAACA	TGGACTTGTC	AAATTTACCT	3600
	GGGGGAATGT	ATCTGAAGTT	AATCGCGAAC	TCGGTGTCTT	TGTTATCAAA	CCATCAGGCG	3660
	TGGATTATGA	CGAATTGACA	CCTGAAAACA	TGGTAGTGAC	TGATCTAGAT	GGAAGATCC	3720
20	CCGGT						3725

## (2) INFORMATION FOR SEQ ID NO:21:

25	(i) SEQUENCE CHARACTERISTICS:						
	(A) LENGTH: 2483 base pairs						
	(B) TYPE: nucleic acid						
	(C) STRANDEDNESS: single						
	(D) TOPOLOGY: linear						
30	(ii) MOLECULE TYPE: DNA (genomic)						
	(iii) HYPOTHETICAL: NO						
35	(iv) ANTI-SENSE: NO						
40	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:						
	TCTAGAAATCA	TTTCCACGCA	GTGGCTCAG	GAAGTCGCAA	TTATCTGGGT	GAGTTTTCAG	60
	CGTGTGGGAC	TGAAGTGGAG	ATGATGAGA	TATACCGCGC	CTTGTGCATG	GCACATTICA	120
45	AGAGTTCGCG	TCCAGATGCC	CAGACCTTGA	TTTTCTATAC	CCACTATGAC	ACTGTGCCAG	180
	CGGATGGGGA	TCAGGCTCTG	ACAGAGGATC	CTTTTACGCT	TTCCGCTCCG	AATGGCTCAT	240
50	GTATGGCGGT	GGGGTTGATG	ACGACAGGGT	CATATCACAG	CTCGCTTGAG	TGCTTGAGAA	300
	AATATATGCA	GCCCTGATGA	TTACCTGTCA	ATATCAGCTT	TATCATGGAG	GGAGCGGAGG	360
	AATCGGCTTC	AACAGACCTA	GATAAGTATT	TGGAAGGCA	TGCAGACAAA	CTCCGTGGGG	420
55	CGGATTGTGT	GGTCTGGGAA	CAAGGGACCA	AAAATGCCTT	GGAACAGCTG	GAAATTTCTG	480
	GTGGCAATAA	GGGGATTGTG	ACCTTTGATG	CCAAGGTAAA	AAGCGCTGAT	GTGGATATCC	540
60	ACTCGAGTTA	TGGTGGTGT	GTGGAATCAG	CTCCTTGGTA	TCTCTCTCAA	GCCTTACAGT	600
	CTCTTCGTGC	TGCGGATGGC	CGTATCTTGG	TTGAAGGCTT	GTACGAAGAA	GTACAAGAGC	660

WO 98/26072

PCT/US97/22578

- 62 -

	CCAATGAACG AGAAATGGCC TTGCTAGAAA CTTATGGTCA ACGAAACCCA GAGGAAGTTA	720
	GTGGATTATA TGGATTGGAG TTGCCTCTCT TACAGGAGGA GCGGATGGCC TTTCTAA AAC	780
5	GTTCCTTTT CGAGCCAGCG CTTAATATCG AAGGAATCCA GTCTGTTAT CAAGSTCAGG	840
	GTGTTAAGAC TATTTTGCTT GCAGAACCCA GTGCCAAGCT AGAGSTTCGT CTGGTTCGGG	900
10	GCCTAGAACG GCATGATGTT CTGGAAAAA TTGGAAACA GCTAGACAAA AATGGCTTTG	960
	ATAAGGTAGA ATTACTAT ACCTTGGGAG AGATACTAGA GTCGAAGCGA TATGAGCGCA	1020
15	CCAGCCATTC TCAATGTGAT CGAGTTGGCC AAGAAATTCT ATCCACAGGG CGTTTCAGTC	1080
	TTGCCGACGA CAGCGGGGAC AGGACCTATG CATACGGTCT TTGATGCCCT AGAGGTACCA	1140
	ATGGTTGCAT TCGGTCTAGG AAATGCCAAT AGCCGAGACC ACGGTGGAGA TGAAGATGTG	1200
20	CGAATCGCTG ATTATTACAC CCATATCGAA TTAGTAGAGG AGCTGATTAG AAGCTATGAG	1260
	TAGAGATATT ATCAAGTTAG ATCAGATCGA TGTGACTTTT CACCAAAAAG AGAGAACCAT	1320
	CACAGCGGTT AAGGATGTGA CCATTACAT CCAAGAAGGG GATATCTACG GAATCGTTGG	1380
25	ATATTCTGGA GCAGGGAAT CAACCTTGT ACGGGTGATT AACCTCTTGC AAAAACCATC	1440
	TGCAGGGAAA ATTACCATTG ACGACGATGT GATTTTTGAC GGCAAGGTGA CCTTGACGGC	1500
30	AGAGCAGTTG CGTCGTAAAC GTCAAGATAT CGGGATGATT TTCCAGCATT TTAACCTGAT	1560
	GAGCCAAAAG ACAGCAGAGG AGAATGTAGC CTTTGGCCTT AAACACTCTG GACTCAGCAA	1620
35	GGAAAGAAAAG AAGGCTAAAG TAGCTAAGTT GTTGACTTTG GTTGGTTTGG CAGATCGTGC	1680
	TGAAACTAC CCTTCACAAC TATCTGGAGG GCAAAAACAG CGTGTGGCAA TTGCGCGTGC	1740
	CTTGCCCAAT GATCCAAAA TCTTGATTTC AGACGAGTCA ACTTCTGCC TTGACCTTAA	1800
40	GACAACCAAG CAGATTTTGG CCTTGTGCA AGATTTGAAC CAAAAATTAG GATTGACAGT	1860
	TGCTTCTATT ACGCATGAAA TGAGATTGTC AAAGACATTG CCAACCGTGT GCGGGTTATG	1920
45	CAGGATGGGC ATTTGATTGA AGAGGGAAAT GTCTTGAAA TCTTCTCAA CCGTAAACAA	1980
	CCTTTGACTC AAGACTTTAT CTCAACAGCC ACAGGTATTG ACGAAGCCAT GGTCAAAATC	2040
	GAGAAGCAAG AAATCGTGA ACACTTGTCT GAAAACAGTC TCTTGGTGA ACTTCAAGTA	2100
50	CGCTGGAGCT TCAACAGACG AGCCACTTTT GAATGAATTG TACAAGCATT ACCAAGTAAT	2160
	GGCTAATATT CTCTATGGGA ATATCGAAAT TCTCGATGGT ACTCTGTTG GAGGAATTGG	2220
55	TGGTGGTCTT GTCAGGTGAA AAAGCAGCGT TGGCAGGTGC CCAAGAGCC ATTCTGCAAG	2280
	CAGGTGTACA ACTAAAAGTA TTGAAGGGAG TACAGTAAGA TGGAAATCAT GATTCAAACC	2340
	TATTACCAA ATGCTATAA AATGGGTTGG GCTGTCAGGC AGGCTGGGG ACGGCTATCT	2400
60	ACTTAACTCT TTATATGCAG TTCTTTCCTT CATTATTCGG GTTCTGGGG CTAGTGGCAG	2460

-63-

GTCTTCTCGT CTTAAGCGCC AGT

2483

## (2) INFORMATION FOR SEQ ID NO:22:

5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1010 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

15 (iv) ANTI-SENSE: NO

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCAATTAATG TGAGTTAGCT CACTCATTAG GCACCCAGG CTTTACACTT TATGCTTCGG	60
GTCGATATGT TGTGTGGAAT TGTGAGCGGA TAACAATTTT ACACAGGAAA CAGCTATGAC	120
25 CATGATTACG CCAAGCTTGC ATGCCTGCAG GTCGACTCTA GAGGATCCAA GCCATAGTTA	180
GACATGACTG CCAAACTTAA GGTTTGAGCA GTTGTTAAAT AAGCATTAGC TGTCGCCTCC	240
30 ATGTTGGGAC TGTTACTTT GAGGCCTACT AAGGCTAGAG ATCCCAACAT CATCAGGATC	300
AAGATGGATA AAAACGCCC TTGAGCCTG TGAAGGACTG AATTAAAGTC TTCCGAATAA	360
35 GTTTTTCGCT TGATCATGCT AGTACTCCAA ACTGTCAATA TCCTGAGGAT GCTGTTGAG	420
CACCACATCC TTGACACTGG CATCGTGCAAT TTGAATCAGC CGATCAGCAA TGGGCGCCAA	480
AGCTCCATTA TGAGTCAGCA TGATCACCCT CGCTCCCTTT TGACGAGACA TGTCTTGGAG	540
40 AATTTTCAAA ACCTGCTTGC CGTCTGATA ATCCAAGGCT CCAGTCGGTT CATCAAAAG	600
GAGAATTTTA GGATTTTGG CTACCGCGCG TGCAATGGAG ACTCGCTGTT GCTCCCTCC	660
AGAAAGCTGG GCTGGAAGT TATTTAGACG ATGAGCCAGA CCTACATCTG TCAAGACCTG	720
45 ATCAGAATTC AAGGCATCTG TCACAATTTT AGAAGCAGTT CCACATTTTC CTTAGCTGTC	780
AGATTAGAAA CTAGATTATA AAACGAAAA ACAACCCCA CATCATTTCT ACGGTAATTG	840
50 GTGCGCTGTT GGGAACTATA ATCCGCAATA TTAACACCAT CAATCCAGAT TTCCCCTTCA	900
TCATTGGTAT CCAATCCCCC AAGAAGGTTA AGAACTGTTG ACTTGCCTGC ACCTGAAGCA	960
55 CCAAGGATAA TAACAGTTT CCCCCTTTCA ATCTCAAAAT TCACATCAGC	1010

## (2) INFORMATION FOR SEQ ID NO:23:

60 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1299 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single

- 64 -

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TCGATCGCAC CGTCTCTCC TCGTCTGCT CTGCTGGGC TATAGTCCC TCTTCTAGTC 60  
 15 TTGATTTTCT TTGCCCATGC GTTCTTACCA CTTCTACTGT TTGCAGGTTT TACATGCTCG 120  
 GATATACTAT TTGTGCTAGG CTAGCTTCT AGGATGGAGA AAAGAAGCT AGTAGAGTTA 180  
 20 TTGAAAGGGG GCATCTTATG ATTGAGTTGA AAAATATTAC CAAAACCAT GGGGAAAAG 240  
 TGATTTTGGG TAACTTATCT CTCAGGATTG ATCAGGGGGA TTTGGTAGCT ATTGTTGGTA 300  
 AGAGTGGTAG TGGGAAGTCG ACCTTGTTAA ATTTATTGGG TTTGATAGAT GGTGATTATA 360  
 25 GCGGACGGTA TGAGATTTT GTTCAGACAA ATCTAGCGGT TAATTCTGCT AAGTCGCAAA 420  
 CAATAATCCG TGAACATATC TCTTATCTGT TTCAAAATTT TGCCCTGATT GATGATGAAA 480  
 30 CGGTCGAGTA CAATCTCATG CTGGCGCTGA AATATGTGAA ATTGCCTAAG AAAGACAAGC 540  
 TCAAAAAGGT GGAAGAGATT TTAGAGAGAG TAGGTTTGTC AGCTACTTTG CATCAAAGGG 600  
 TCTCCGAGTT GTCTGGGGGC GAACAACAAC GAATTGCAGT TGCTAGAGCC ATCTTAAAC 660  
 35 CCAGCCAGCT GATTTTAGCC GATGAACCTA CAGGTTCCGT GGATCCTGAA AATGAGATT 720  
 TGGTCTTGAA GTTCTCTTA GAGATGAATC GAGAAGGGAA AACAGTCATT ATTGTGACCC 780  
 40 ACGATGCTTA TGTAGCCCAA CAATGTATC GTGTATTGA ATTGGGCGAG GGAATATGAG 840  
 TTCATTGAGC TCCTTTTGAC TGGCTGAATA CTCATGTTTT CCAGAGAAAA ATAGCATAAA 900  
 TAGCGCTAGG AATGACATTT TATGTAGCAT TTCTAGGTTT TTTTGTTTCA AATTGAAAAT 960  
 45 TTTTTCATTT TAGGCTTGAC AAAGGATGAG TATAGGAGTA TTATTATAC AATAAAAAAG 1020  
 AATAACATA AAGAAGGCTT TGTATGAAT AAGATGAAGA AGGTGTTGAT GACGATGTTT 1080  
 50 GGTTTAGTGA TGCTCCCCCT ACTATTGCT TGTAGTAA CAATATCGGC TGGAAATTGAA 1140  
 GCCATCAAGT CCAAGGAAA ATTGGTTGTA GCCCTCAATC CAGATTTTGC TCCATTGAA 1200  
 55 TATCAAAAAG TGTTGATGG GAAAAATCAG ATTGTGGGTT CAGATATCGA CTTAGCCAAG 1260  
 CTATCGCAAC AGAAGTAGGT GTCGACTCTA GAGGATCCC 1299

(2) INFORMATION FOR SEQ ID NO:24:

60 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 252 base pairs

WO 98/26072

PCT/US97/22578

- 65 -

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

15	GCCAAACCAA AACAACTCTC AGGTGGTCAA AAACAACGTG TGGCCATCGC TCSTGCCCTC	60
	TCCATGAATC GCGAGGCTAT TCTCTTTGAT GAACCAACAT CAGCTCTCGA TCCAGAAATG	120
20	GTTGGAGAAG TCCTCAAAAT CATGCAGGAC CTGGCTCAGG AAGGCTTGAC CATGATTGTC	180
	GTAACCCATG AAATGGAATT TGCCCGTGAT GTCTCTCACC GTGTTATCTT TATGGATAAG	240
	GGCCTGATCC CC	252

25

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 455 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

30

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

45	GATTGGTAC CTCTTCGCAA GGAAGTCGGC ATGGTTTTTC AACATTTTAA CCTTTATCCA	60
	CACAAAACGG TGTTAGAAAA CGTGACACTT GCGCCCATTA AAGTCTAGG AATTGATAAA	120
50	AAAGAAGCTG AAAAAACCGC CCAAAAATAT CTGGAATTTG TAAATATGTG AGACAAGAAA	180
	GATTCCTATC CCGCCATGCT ATCTGGTGGG CAAAAACAGC GGATCGCCAT CGCTCGTGGT	240
	CTTGCTATGC ATCCGGAACT CTTCTCTTTT GATGAACCAA CATCTGCTCT TGATCCTGAG	300
55	ACTATCGGAG ATGTTCTAGC AGTTATGCAG AAAGTGGCGC ATGATGGGAT GAACATGATC	360
	ATCGTTACCC ACGAAATGGG CTTTGCTCGA GAGGTTGCGG ACCGCATTAT CTTTATGGCC	420
	GACGGAGAAG TTTTAGTAGA TACGACAGAT GTCGA	455

60

(2) INFORMATION FOR SEQ ID NO:26:

- 66 -

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 913 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID No:26:

20	AAACACTGCT TCTTGAGCGA ATGACGCTTT GTCCTTTTAA TGAGGTTACC AACGGCTTCA	60
	AAGAGGATTC CCAGCTCGTT CAGCTGTGGA GGTAGCTCGT CTTCCTCGTG ATGTAAAAGT	120
	CGAAATTGAA GTATCCGCGAG AGATTGGATA AGCTAGTTGA AGTTTGGTGT TGCCAAACTT	180
25	CTTTTGATAT AAGGAGAAAA AGATGACAAA GAAACAATT CACTTGGTGA TTGTGACAGG	240
	GATGGGTGGC GCAGGAAAA CTGTAGCCAT TCAGTCCTTC GAGGATCTAG GTTATTTCAC	300
	CATTGATAAT ATGCCGCCAG CTCTCTTGCC TAAGTTTTTG CAGCTGGTTG AAATTAAGGA	360
30	AGACAATCCT AAGTTGGCCT TGGTAGTGA TATGCGTAGT CGTTCCTTCT TTTCAGAGAT	420
	TCAAGCTGTT TTGGATGAGT TGGAAAATCA AGATGGTTTG GATTTCAAAA TCCTCTTTTT	480
35	GGATGCGGCT GATAAGGAAT TGGTCGCTCG TTACAAGGAA ACCAGACGGA GTCACCCACT	540
	AGCAGCAGAC GGTCTGATTT TAGATGGAAT CAAGTTGGAA CGTGAACCTC TGGCACCTTT	600
40	GAATAATATG AGCCAAATG TGGTGGATAC GACTGAACCT ACTCCACGTG AGCTGCGCAA	660
	AACCCCTTGA GAGCAGTTTT CAGACCAAGA ACAAGCTCAG TCTTTCCGTA TCGAAGTCAT	720
	GTCTTTCGGA TTTAAGTATG GAATCCCGAT TGATGCGGAC TTGCTCTTTG ATGTCCGTTT	780
45	CTTGCCAAAT CCCTATTATT TACCAGAACT GAGAAACCAA ACGGGTGTGG ATGAACCTGT	840
	TTATGATTAT GTCATGAACC ATCTGAGTC AGAAGACTTT TATCAACATT TATTGGCCTT	900
50	GATTGAGCCG ATT	913

(2) INFORMATION FOR SEQ ID No:27:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5919 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO



(iv) ANTI-SENSE: NO

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

	TCGATTCTGTT	60
10	GGTTTGAAT TGTATTGCA GTTATTAGAG GAAGCTATTG CTAACGAAA CGGTAATGCT	120
	AACGCTAACA CAAGAACCAA AGGGAATGCT GAGTTGATTT TGCAAATTGA TGCCTATCTT	180
15	CCTGATACTT ATATTTCTGA TCAACGACAT AAGATTGAAA TTACAAGAA AATTCGTCAA	240
	ATTGACAACC GTGTCAATTA TGAAGAGTTA CAAGAGGAGT TGATAGACCG TTTTGGAGAA	300
20	TACCCAGATG TAGTAGCCTA TCTTTTAGAG ATTGGTTTGG TCAATCATA CTTGGACAAG	360
	GTCTTTGTTT AACGTGTGGA AAGAAAAGAT AATAAAATTA CAATTCAAAT TGA AAAAGTC	420
	ACTCAACGAC TGTTTTATG TCAAGATTAT TTTAAAGCTT TATCCGTAAC GAACCTAAAA	480
25	CGAGGCATCG CTGAGAATAA GGGATTAAAT GAGCTTGAT TTGATGTCCA AAATAAGAAA	540
	GATTATGAAA TTTTGAAGG TCTGCTGATT TTTGGAGAAA GTTTATTAGA GATAAAAGAG	600
	TCTAAGGAAA AAAATTCCAT TTGATATTTT TCTTCTATAA AATAGATAAA ATGGTACAAT	660
30	AATAAATTGA GGTAATAAGG ATGAGATTAG ATAAATATTT AAAAGTATCG CGAATTATCA	720
	AGCGTCGTAC AGTCGCAAAAG GAAGTAGCAG ATAAAGGTAG AATCAAGGTT AATGGAACTC	780
35	TGGCCAAAAG TTCAACGGAC TTGAAAGTTA ATGACCAAGT GAAATCGCTT GGCAATAAGT	840
	TGCTGCTTGT AAAGGTACTA GAGATGAAAG ATAGTACAAA AAAAGAAGAT GCAGCAGGAA	900
	TGTATGAAAT TATCAGTGAA ACACGGGTAG AAGAAAATGT CTAAAAATAT TGTACAATTG	960
40	AATAATCTTT TTATTCAAA TGAATACCAA CGTCGTCGCT ACCTGATGAA AGAACGACAA	1020
	AAACGGAATC GTTTTATGGG AGGGGTATTG ATTTTGATTA TGCTATTATT TATCTTGCCA	1080
45	ACTTTTAATT TAGCGCAGAG TTATCAGCAA TTAATCCAAA GACGTCAGCA ATTAGCAGAC	1140
	TTGCAACTC AGTATCAAAC TTTGAGTGAT GAAAAGGATA AGGAGACAGC ATTTGCTACC	1200
	AAGTTGAAAG ATGAGAGTTA TGCTGCTAAA TATACACGAG CGAAGTACTA TTATTCTAAG	1260
50	TCGAGGGAAA AAGTTTATAC GATTCTGAC TTGCTTCAA GGTGATAAAA TGGAAAATTT	1320
	ATTAGACGTA ATAGAGCAAT TTTTGAGTTT GTCAGATGAA AAGCTGGAAG AATTGGCTGA	1380
55	TAAAAATCAA TTATTGCGTT TACAAGAAGA AAAGGAAAGG AAGATGCGT AAATCTTAA	1440
	TTATTTTGTT GCTACCAAGT TTTTGGACCA TTTCAAAAGT CGTTAGCACA GAAAAGAG	1500
	TGCTCTATAC TTGCAAGGAA ATTTATTACC TTTCAAAATC TGACTTTGGT ATTTATTTTA	1560
60	GAGAAAAATT AAGTTCTCCC ATGGTTTATG GAGAGTTTCC TGTTTATGCG AATGAAGATT	1620

PCT/US97/22578

	TAGTAGTGGA	ATCTGGGAAA	TTGACTCCCA	AAACAAGTTT	TCAAAATACC	GAGTGGCGCT	1680
5	TAAATAACAA	AGGAATCCCA	GTATTTAAGC	TATCAAAATCA	TCAATTATATA	GCTGCGGACA	1740
	AACGATTTTT	ATATGATCAA	TCAGAGGTAA	CTCCACAAT	AAAAAAAAGTA	TGGTTAGAAT	1800
	CTGACTTTAA	ACTGTACAAT	AGTCCTTATG	ATTTAAAAGA	AGTGAAAATCA	TCCTTATCAG	1860
10	CTTATTCCGA	AGTATCAACT	GACAAGACCA	TGTTTGTAGA	AGGAAGAGAA	TTTCTACATA	1920
	TAGTACAGCG	TGGATGGGTA	GCTAAAGATC	CAACTCTCTGA	AGAAGATCAAT	CGGATGAGTA	1980
15	TTGTTCAGAA	AATGTATATC	AAAAAATATC	AGAAAGTCTA	TTTCTCTPAT	TATGTTAAGC	2040
	AACCTACTAC	TGGAAGAAAG	GCTGGTATCA	TATCAAGATGA	AAAGATGTAT	GACGCCAGCG	2100
	TTTTGAAACT	CTCTTATCTC	TATTATACGC	AAGAAAAAAA	TAAATGAGGG	TCTTTATCAG	2160
20	TTAGATACGA	CTGTAAAAAT	CGTATCTGCA	GTCAATGATT	TTCCAGGTTCT	TTATAAACCA	2220
	GAGGGAAGTG	GTAGTCTTCC	TAAAAAGAA	GATAATAAAG	AATATTCTTT	AAAGGATTTA	2280
	ATTACGAAGG	TATCAAAGAA	ATCTGATAAAT	GTAGCTCATATA	ATCTATTGGG	ATATTACATT	2340
25	TCAAACCAAT	CTGATGCCAC	ATTCAAAATC	AAGATGTCTG	CCATTATGGG	AGATGATTGG	2400
	GATCCAAAAG	AAAAATTGAT	TTCTTCTAAG	ATGGCCGGGA	AGTTTATGGA	AGCTATTTAT	2460
30	AATCAAATGT	GATTTGTGCT	AGAGTCTTTG	ACTAAAACAG	ATTTTGATAG	TCAGCGCAATT	2520
	GCCAAAGGTG	TTTCTGTAA	AGTAGCTCAT	AAAATTGGAG	ATGCGGATGG	ATTTAAGCAT	2580
35	GATACGGGTG	TTGTCTTCTA	AGATCCTCCA	TTTATTTCTT	CTATTTTTCA	TAAGAATTTCT	2640
	GATTATGATA	CGATTTCATG	GATACGCCAG	GATGTTTATG	AGSTTCTAAC	ATGAGGGAAAC	2700
	CAGATTTTTT	AAATCAATTT	CTCAAGGAAG	GATATTTCAA	AAGCATGTGT	AAGGCGGTTCT	2760
40	TAGTCTTTTC	TGGTGGGATTA	GATTCCATGT	TTCTATTTAA	GGTATTGTCT	ACTTATCAAA	2820
	AAGAGTTAGA	GATTGAATTG	ATTCTAGTCT	ATGTGAATCA	TAAAGCAGAA	ATTGAATCAG	2880
45	ATTGGGAAGA	AAAGGAATTA	AGGAAGTTGG	CTGCTGAAGC	AGAGCTTCTCT	ATTTATATCA	2940
	GCAATTTTTCT	AGGAGAATTT	TCAGAAGCGC	GTGCACGAAA	TTTTCGTTAT	GATTTTTTTTC	3000
	AAGAGGTCCA	TGAAAAAGAC	AGGTGCGACA	GCTTTAGTCA	CTGCCACCA	TGCTGATGAT	3060
50	CAGGTGGAAA	CGATTTTTAT	GCGCTTGATT	CGAGGAACCT	CCTTGCCTTA	TCTATCAGGA	3120
	ATTAAAGAGA	AGCAAGTAGT	CGGAGAGATA	GAATCATCTC	GTCCCTTCTT	GCATTTTCAG	3180
55	AAAAAGACT	TTCCATCATAT	TTTTCTACTTT	GAAGATACAT	CAAAATCAGGA	GAATCATTTAT	3240
	TTTCCGAATC	GTATTCGAAA	TTCTTACTTA	CCAGAAATGAG	AAAAAGAAA	TCCTTCGATT	3300
	AGGATGTCAA	TCTTTAGCCA	TGGCAAGATA	AAATTTAGAT	TATGATTTGG	CAATAGCTGA	3360
60	ATTATCTAAC	AATATTAAATG	TGGAAGATTT	ACAGCAGTTA	TTTTCTTACT	CTGAGCTTAC	3420

WO 98/26072

PCT/US97/22578

- 69 -

	ACAAAGAGTTT	TTACTTCAAA	CTTATCTGAA	TGTTTTTCCA	GATTTGAATC	TTACAAAAAGC	3480
	TCAGTTTGGT	GAAGTTTCAG	AGATTTTAAA	ATTTAAAAAGC	CAGTATCGTC	ATCCGATTAA	3540
5	AAATGGCTAT	GAATTGATAA	AAGAGTACCA	ACAGTTTCAG	ATTGTGAAAA	TCAGTCCGCA	3600
	GGCTGATGAA	AAGGAAGATG	AACCTGTGTT	ACACTATCAA	AATCAGGTAG	CTTATCAAGG	3660
	ATATTTATTT	TCTTTTGGAC	TTCCATTAGA	AGGTGAATTA	ATTCACAAAA	TACCTGTTTC	3720
10	ACGTGAAACA	TCCATACACA	TTGCTCATCG	AAAAACAGGA	GATGTTTTCG	TTAAAAATGG	3780
	GCATAGAAAA	AACTCAGACG	GTTTATTTAT	TGATTTGAAA	ATCCCTATGG	AAAAGAGAAA	3840
15	CTCTGCTCTT	ATTATTGAGC	AATTTGGTGA	AATTGCTCTA	ATTTTGGGAA	TGCGACCAA	3900
	TAATTTGAGT	AAAAAAACGA	AAAATGATAT	AATGAACACT	GTACTTTTATA	TAGAAAAAAT	3960
	AGATAGGTAA	AAAATGTTAG	AAAACGATAT	TAAAAAAGTC	CTCGTTTCAC	ACGATGAAAT	4020
20	TACAGAAGCA	GCTAAAAAAC	TAGGTGCTCA	ATTAACATAA	GACTATGCAG	GAIAAAATCC	4080
	AATCTTAGTT	GGGATTTTAA	AAGGATCTAT	TCCTTTTATG	GCTGAATTGG	TCAAACATAT	4140
25	TGATACACAT	ATTGAAATGG	ACTTCATGAT	GGTTTCTPAG	TACCATGGTG	GAACAGCAAG	4200
	TAGTGGTGTT	ATCAATATTA	AACAAGATGT	GACTCAAGAT	ATCAAAGGAA	GACATGTTCT	4260
	ATTTGTAGAA	GATATCATTTG	ATACAGGTCA	AACCTTGAAG	AATTTGCGAG	ATATGTTTAA	4320
30	AGAAAGAGAA	GCAGCTTCTG	TTAAAATTGC	AACCTTGTGG	GATAAACCCAG	AAGGACGTGT	4380
	TGTAGAAATT	GAGGCAGACT	ATACCTGCTT	TACTATCCCA	AATGAGTTTG	TAGTAGGTTA	4440
35	TGGTTTAGAC	TACAAAGAAA	ATTATCGTAA	TCTTCCTTAT	ATTGGAGTAT	TGAAGAGGGA	4500
	AGTGTATTCA	AATTAGAAAG	AATAATCTTT	AATGAAAAAA	CAAAATAATG	GTTTAATTAA	4560
	AAATCCTTTT	CTATGGTTAT	TATTTATCTT	TTTCCTTGTG	ACAGGATTCC	AGTATTTCTT	4620
40	ATTCTGGGAA	TAACTCAGGA	GGAAGTCAGC	AAATCAACTA	TACTGAGTTG	GTACAAGAAA	4680
	TTACCGATGG	TAAATGAAAA	GAATTAACCT	ACCAACCCAA	TGTTAGTGTG	TCGAAGTTTC	4740
45	TGTTGTCTAT	AAAAATCCTA	AAACAAGTAA	AGAAGGAACA	GGTATTCAGT	TTTTACGCC	4800
	ATCTGTACTT	AAGTAGAGAA	AATTTACCAG	CACTATTCTT	CCTGCAGATA	CTACCGTATC	4860
	AGAATTGCAG	AACTTGCCTA	CTGACCATAA	AGCAGAAGTA	ACTGTTAAGC	ATGAAGTTTC	4920
50	TAGTGGTATA	TGGATTAAATC	TACTCGTATC	CATTGTGCGA	TTTGGAAATC	TATTCCTCTT	4980
	CCTATTCTCT	ATGATGGGAA	ATATGGGAGG	AGGCAATGGC	CGTAATCCAA	TGAGTTTGG	5040
55	ACGTAGTAAG	GCTAAAGCAG	CAAATAAAGA	AGATATTAAA	GTAAGATTTT	CAGATGTTGC	5100
	TGGAGCTGAG	GAGAAAAAAC	AAGAACTAGT	TGAAGTTGTT	GAGTCTTAA	AAGATCCAAA	5160
	ACGATTACCA	AAACTGGGAG	CCCGTATTCC	AGCAGGTGTT	CTTTTGAGG	GACCTCGGG	5220
60	GACAGGTAA	ACTTTGCTTG	CTAAGGCAGT	CGCTGGAGAA	GCAGGTGTTG	CATTCTTTAG	5280

WO 98/26072

PCT/US97/22578

-70-

TATCTCAGGT TCTGACTTTG TAGAAATGTT TGTCGGAGTT GGAGCTAGTC GTGTTCCGTC 5340  
 TCTTTTTAG GATGCCAAAA AAGCAGCACC AGCTATCATC TTTATCGACT GAAATGGATG 5400  
 5 CCCGTGGGAC GTCAACGTGG AGTCGGTCTC GCGGAGGTA ATGACGAACG TGAACAAACC 5460  
 TTGAACCAAC TTTTGATTGA GATGGATGGT TTTGAGGGA ATGAAGGGAT TATCGTCATC 5520  
 10 GCTGCGACAA ACCGTCAGA TGTACTTGAT CCGCCCTTT TCGCTCCAGG ACGTTTGAT 5580  
 AGAAAAGTAT TGGTTGGCCG TCCTGATGTT AAAGTCGTG AAGCAATCTT GAAAGTTCAC 5640  
 GCTAAGAACA AGCCTTTAGC AGAAGATGTT GATTGAAAT TAGTGGCTCA ACAAACTCCA 5700  
 15 GGCTTTGTTG GTGCTGATTT AGAGAATGTC TTGAATGAAG CAGCTTTAGT TGCTGCTCGT 5760  
 CGCAATAAAT CGATAATTGA TGCTTCAGAT ATGATGAAG CAGAAGATAG AGTTATTGCT 5820  
 20 GGACCTTCTA AGAAAGATAA GACAGTTTCA CAAAAAGAAC GAGAAATGGT TGCTTACCAT 5880  
 GAGGCAGGAC ATACCATTGT TGGTCTAGTC TTGTCGACT 5919  
 (2) INFORMATION FOR SEQ ID NO:28:  
 25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1863 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 30 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: DNA (genomic)  
 (iii) HYPOTHETICAL: NO  
 35 (iv) ANTI-SENSE: NO  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:  
 40 GAGCTCGGTA CCCGGGGATC ATACTCAAGA GGAGGTAATC CAATGAACAC TAGCTTAAA 60  
 45 CTCAGCAAC AACTCAGTTT TGGAGAGGAG ATTGCTAATA GCGTGACCCA TGCTGTGGGT 120  
 GCAGTCATCA TGCTTATCTT GCTGCCTATT TCATCCATCT ATAGTTATGA AGCACACGGA 180  
 50 TTTTATCCT CTATCGGCGT TTCCATTTTC GTCATCAGTC TCTTTCTCAT GTTCTCTATCA 240  
 TCCACCAATT ATCACTCTAT GGCCTATGGT TCGACCCACA AATATGTTT GCGAATCATT 300  
 GACCATTCTA TGATTTACGT TGCCATTGCC GGCTCATACA CGCCCGTTGT CTTGACCTTG 360  
 55 ATGAATAACT GGTITGGCTA TCTGATTATT GTCATCCAAT GGGGAACGAC CATCTTGGT 420  
 ATTCTCTATA AAATCTTTGC TAAAAAGGTC AATGAGAAAT TTAGCCTTGC TCTTTACCTG 480  
 ATTATGGGCT GGTITGGTCT GGCTATCATT CCGGCCATTA TCAGTCAAA GACACCGGTT 540  
 60 TTCTGGAGTC TCATGGTAAC TGGCGGAGTC TGTATACAG TTGAGCTGG ATTTTATGCC 600

PCT/US97/22578

5	AAGAAAAAC	CTATTTC	CATGATTGG	CATCTCTTTA	TCCTAGCTGC	GTCCGCACTC	660
	CAATACATCG	CTATTGTTT	TTACATGTAA	AAAAGTTGAG	AAATTCAACT	TCAACTTTTT	720
	TCTTTACACA	TATTGATAAA	GTACTGGTGC	AAGCGCACAT	CATCAGTCAA	TTCTGGATGA	780
10	AAAGAACCTA	CCAACATATT	TTTTCTTTGG	GCTGCAACAA	TTTGATTGTT	CACTATTGCT	840
	AAAATTTCTA	CACCTCACC	AACACTACTG	ATAATCGGAC	CACGGATAAA	GGTCATTGGA	900
	ATCTTGCCAA	CTCCCTTACA	TTCTGCTTCC	GTGTAGAAAC	TTCTTAATTG	CGGCCCATAA	960
15	GCATTACGCT	CGACCACTAT	ATCCATCATT	CCTAGATGAC	TCTCTTTCTG	AGAAAGTAT	1020
	TCTTTAGCCA	GCAAAATTAA	GCCCGCACAG	GTCCCAACAG	CTGGTAAGCC	AGATAGAATG	1080
	GCTTCTCGTA	TGGGAAGTAG	CATGTTCTGG	TCACGTAAAG	GCTTGCCCAT	GGTGTAGAC	1140
20	TCACCACGAG	GCAAAATAAA	CCGACAAGT	CACTCTGATC	TTGCTGAAA	CATCTAGATT	1200
	TCTGAGTCT	ACACTCTCGA	CACCTAATTG	ATCTAGCACT	TTTGCATGTT	CTGCAAAAGC	1260
	CCCTTGCAAG	GCCAAATTTC	CGATTTTCAT	CTATTTTCTC	CGTTCAGCCA	TGAGAATTTG	1320
25	GATTCAATTT	CATTAAATAC	AACCATGGCT	TCTCTTAAAT	CTTCAGAGAT	TTGAGCTAGG	1380
	ATTTGAGGAT	TACGGAGATT	AGTCACAGCC	TTAACAATGG	CACTCGCTCG	TTTAACAGSA	1440
	TCTCTGACT	TGAAAATACC	TGAACCGACA	AAGACCCCTT	CTGCCCTTAA	TTGCATCATT	1500
30	AACGCGACAT	CTGCTGGCGT	TGCAACACCT	CCAGCAGCGA	AATTTACAAC	TGGCAATTTT	1560
	CCATGTTTAT	GAACATATTG	GACCAATTCT	ACAGGGACTT	GCAAAATCCT	GGCAGCAACA	1620
	TAAAGCTCGT	CCTCAGCTAA	GTTTGTGAAT	CGCGGAATTT	CCTGATTCAT	CATACGCATA	1680
35	TGACGAACAG	CTTGAGACTAT	ATCCCTGTGC	CTTGGTCTCT	CTTAGTAGAC	AATCATGGAA	1740
	GCACCTTTCAG	CGATACGACG	CAGCGCTTCA	CCCAAAATCT	TAGCACACCA	GACAAAAGGA	1800
	ACTTGGAATT	CTTTCTTGTG	CACATGGAAA	CGGTATCAGC	CTGAGAGATG	AACCTCACTC	1860
40	TCG						1863

50 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 5448 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

60

PCT/US97/22578

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

5	TAAAGAAGGT	GGATTTGAAG	TTAACGGTAA	ATTCATCAAA	GTTTCTGCTG	AACGTGATCC	60
	AGAACAATCT	GACTCGGGTCA	CTGACGGTGT	AGAAATCGTT	CTTGAAGCTA	CTGCTTCTTT	120
	TGCTAAGAAA	GAAGCAGCTG	AAAAACACCT	TAAAGGTGGA	GCTAAAAAAG	TTGTTATCAC	180
10	TGCTCTCGST	GGAAACGACG	TTAAAACAST	TGTATTCAAC	ACTAACCACG	ACGTTCTTGA	240
	CGGTACTGAA	ACAGTTATCT	CAGGTGCTTC	ATGTACTACA	AACTGCTTGG	CTCCAATGGC	300
	TAAAGCTCTT	CAAGACAATCT	TTGGTGGTGT	TGAAGGATTA	ATGACTACTA	TCCACGCTTA	360
15	CTACTGGTGAC	CAAAATGATCT	TTGACGGGAC	ACACCGCTGTG	GTAACCTTCT	CCGCTGCTGC	420
	GCTGGTGTGCT	CAAACTACTCGT	TCCTAACTCA	CTGGTGGCTG	CGAAACCTAG	CGGCTTTGTGA	480
20	ATCCCAAGAT	TGAATGGTAA	ACTTGATGGA	TCTGCACAAC	CGGTTCCAAC	TCCAAC TGGA	540
	TCAGTTACTG	AATTGGTCGC	AGTTCTTGAA	AAGAACGTTA	CTGTTGATGA	AGTGAACGCA	600
	GCTATGAAAG	CAGCTTCAAA	CGAATCATAC	GGTTACACAG	AAGATCCAAT	CGTATCTTCA	660
25	GATATCGTAG	GTATGTCTTA	CGGTTCAATT	TTTGACGCCA	CTCAAACTAA	AGTTCTTGAC	720
	GTTGACGGTA	AACAATTTGT	TAAAGTTGTA	TCATGGTAGC	ACACAGCAAT	GTCAATACACT	780
30	GCACAACCTG	TTCGTACTCT	TGAATACTTC	GCAAGAGATT	CTAAATAATT	CTTGAGTTGA	840
	TAGAAAGCAA	GGCTTTGTGG	TCTTGCTTTT	TTATATGGAA	AAATGGATGA	CACGATCATC	900
	TACTCTTTTT	TAATTTCTTT	TCAAATGTAT	TTGAAAGGAT	AGTGAAGGTT	AGGCTCTCTA	960
35	AAGCTAAGTG	GTTGGCTTAAG	GAAGATCGTA	AGGCATAGGT	CTAAGCCGGT	CTTGTCTTTG	1020
	ATTTTACTATT	ATAGACGAGG	GTCTCCGAGA	TAGGAAGATT	ATGATGCAAA	AGGTGCAACAC	1080
40	GAATCTGATG	GGTTCGCGCT	GTCTTTAGCT	TGCAATGAGC	CAAGGAAGTC	TTGTTTGAGA	1140
	ATTGCTTTAA	TCTGCTTACA	TGCGTTTCAG	CATATTTCCC	ATTTTTTGCA	TCAACTATTTC	1200
	TTTTCTACG	ATCATGGGCA	TCACGTCCAA	TTTTGTCTCT	GAAGAACAGT	TCTTTTCTGT	1260
45	TGATATTTC	ATCAACTAGA	GCCCAATATT	CTCTAGAAAT	CTCTTTTTC	TCCAATAAGC	1320
	GATTGAGAAAT	GGGCAGGATA	AAAGGATTTT	TGGCAAGAG	AACTAAGCCA	CTGGTTTCCA	1380
50	TGTCCAGACG	ATGAACGACA	TAGCAGGTTT	GGCCAACATA	GGTACTGACA	TGGTTAAGAA	1440
	GGGCAATTTC	GTTTGGTTGA	TTACCATGCG	TTTTCATCCC	CTCTGGTTTG	TTTACAATAA	1500
	TCAAGTGTGG	ATCTTGATAA	ACTTCCTGCA	CTAAGTCTGG	ATTCGCCCAA	GGGATCGTCT	1560
55	TTTGGGGATA	ATCTTCTCTG	TCAAAAAGTA	AACTGCCAA	GTTCTCCGAA	TTTACGATTT	1620
	CGTTTCACGG	GACTTCTCTT	TGATTTATCA	TAAGTTGTTT	CTTGATCTCT	AAAAAATGAC	1680
60	GGATTTTCT	AGGGATGAGG	AGTTGTTCTT	CAAGTAATTG	CTTTACCGTC	ATTTAGAGTA	1740

WO 98/26072

PCT/US97/22578

-73-

	GAGAGGCGGG	TAATGTAAT	GTGAATTGCA	TACAGATATT	GTAAACAAAA	AAGCCCTATT	1800
	TGGATAGGAA	ATAGCTAAAT	TCTTGTCTTC	CTATGATGAA	GATGATAAAA	TAAACGCATG	1860
5	AAATTAGATA	AATTATTGGA	GAAATTTCTT	TCTCTTTTAA	AAAAAGAAAC	AAGTGAACATA	1920
	GAGGACTCTG	ATTCTACTAT	CTTACGCTGC	TCTCGTAGTG	ATCGAAAAAA	ATTAGCCCAA	1980
	GTAGGTCCGA	TTCGAAAAAT	CTGGCGTCGT	TATCATCTAA	CAAAGATTAT	CCTTATACTA	2040
10	GGTTTGAGTG	CAGGCTTGCT	AGTTGGAATC	TATTGTTTGG	CTGTAGCCAA	GTCCGACCAAT	2100
	GTCAATGATT	TGCAAAATGC	CTTGAAAACT	CGGACTCTTA	TTTTTGACCG	TGAAGAAAAA	2160
15	GAGGCTGGTG	CCTTGTCTGG	TCAAAAGGGA	ACCTATGTTG	AGCTGACTGA	CATCAGTAAA	2220
	AACCTGCAGA	ATGCTGTTAT	TGCGACAGAA	GACCGTTCTT	TCTATAAAAA	TGACGGGATT	2280
	AACATGAGCC	GTCTCTTCTT	GGCTATTGTC	ACTGCTGGAC	GTTCAGGTGG	TGGCTCTACC	2340
20	ATTACCCAAC	AGCTGGCTAA	AAACGCCTAT	TTATCGCAGG	ATCAAACTGT	TGAGAGAAAA	2400
	GCGAAAGAA	TTTTCTTTGC	CTTAGAATTA	AGCAAAAAAT	ATAGTAAGGA	GCAAAATCTA	2460
25	ACCATGTACC	TTACAACAGC	TTAATTTTGA	AATGGTGTGT	GGGGTGTAGA	AGATGCGAGT	2520
	AAGAAATACT	TTGAGTTTTC	TGCATCAGAA	GTGAGTCTGG	ATCAAGCTGC	GACTCTGGCA	2580
	GGGATGCTCA	AGGGGCGGGA	ACTGTATAAT	CCCTTGAATT	CCGTAGAAGA	TTCTACTAAT	2640
30	CGGCGCGATA	CTGTCTTGCA	GAATATGTTT	GCAGCAGGAT	ATAATTGATA	AAACCAAGAA	2700
	ACCGAAGCTG	CTGAAGTTGA	TATGACTTCG	CAATTGCGAG	ATAAGTATGA	AGGAAAAATC	2760
35	TCAGATTACC	GTATCCCTTC	TTATTTTGGT	GCGGTGGTTA	ATGAAGCTGT	TTCCAAGTAT	2820
	AATCTAACAG	AGGAAGAGAT	TGTCATAAAT	GGCTACCGCA	TTTACACAGA	GCTGGACCAA	2880
	AACCTACCAAG	CAATATATGCA	GATTGTTTAT	GAACACACAT	CGCTATTTCG	GAGGGCAGAG	2940
40	GATGGAACGT	TTGCTCAATC	AGGAAGTGTA	GCTCTCGAAC	CGAAAAACAG	GGGAGTTCTG	3000
	GGAGTTGTCT	GTCAAGTTGC	TGACAATGAT	AAAACCTGGT	TCCGGAATTT	CAACTATGCA	3060
45	ACCCAATCAA	AGCGTAGTCC	TGGTTCTAGA	ATTAAGCCTT	TAGTTGTTTA	TACACCAGCA	3120
	GTGGAAGCAG	GCTGGGCTTT	GAATAAGCAG	TTGGATAACC	ATACCATGCA	GTATGATAGC	3180
	TATAAGGTTG	ATAACTATGC	AGGGATCAAA	ACAAGTCGAG	AAGTTCCTAT	GTATCAATCC	3240
50	TTGGCAGAA	CGCTTAATCT	ACCTGCTGTT	GCCACTGTTA	ATGATTTGGG	TGTTGACAAAG	3300
	GCTTTTGAGG	CAGGCGAAAA	ATTGCGACTC	AACATGGAAA	AGGTCGACCG	TGTTCTTGGT	3360
55	GTGCGCTTGG	GAGCGGTGT	TGAAACCAAC	CCTCTTCAAA	TGGCTCAAGC	ATACGCTGCC	3420
	TTTGCAATAG	AAGGTTTAA	GCCTGAAGCT	CATTTTATTA	GTAGAATTGA	AAATGCTAGT	3480
	GGACAAGTTA	TTGCGAGTCA	TAAAAATTCA	CAAAAACGGG	TGATTGATAA	GTCTGTAGCT	3540
60	GACAAGATGA	CCAATATGAT	GTGGGGGACT	TTCAACCAAG	TACCCGGTAT	TAGTTCATCG	3600

WO 98/26072

PCT/US97/22578

-74-

	CCTGCAGACT ATGTCATGGC AGGGAJAACT GGAACAACTG AAGCAGTTTT CAATCCGAG	3660
	TACACAAGTG ACCATGGGT AATTGGTTAT ACTCCGGATG TAGTGATTAG CCACTGGGTT	3720
5	GGCTTCCGA CCACTAGTGA AAATCACTAT CTAGCTGGCT CTACTTCAAA CGGTGCACT	3780
	CATGCTTTTA GAAACATTGC CAATACTATT TTACCTTATA CCCCAGGAAG TACCTTTACG	3840
10	GTTGAAAATG CTTATAAGCA AAATGGAATT GCACCAAGCCA ATACAAAAAG ACAAGTACAA	3900
	ACCAATGATA ATAGCCAGAC AGATGATAAT TTGCTGATA TTCGAGGGCG TGCACAAAGT	3960
	CTAGTAGATG AGGCTAGCCG GGCTATCTCA GATGCGAAGA TTAAGGAAAA GGCTCAJACA	4020
15	ATATGGGATT CGATAGTCAA TCTATTTTCG TAAGATGCTT GTCAAAGCCT AGCTTTCTTG	4080
	TTATAATGGA TAAGATGGAG GCGTTATGGC ACTAAAAAAA GCAAGCCTAG CTTGTGCGGT	4140
20	TTGTGGTTCG AGAACTATT CAATCAAGAT CAGCGGAAAC CCCAAGCCTA CACGACTAGA	4200
	AGTAAATAAA TTTTGTAAAG ATTGTGGCAA GTACACTACA CACAGAGAAA CGAGATAGGA	4260
	GAGAGCGATG CGTTTTATTG GAGATATTTT TAGACTTCTT AAAGACACAA CATGGCCAAC	4320
25	TCGCAAGGAA AGCTGGAGAG ATTTTCGTTT TATCATGGAA TACACAGCTT TCTTTGATG	4380
	AATTATTTAC ATTTTGTACC AGTTGATTGT TTCAGTTTG ATTGATTTA TTAACATTTT	4440
30	TTAGAAGATT AGTGGAGTTA ATTACACTAG AAATCTTCTA TTTATGAAG GAAATATCAT	4500
	GGATAGTTTT GATAAAGGAT GGTTTGTTTT ACAAACTTAT TCTGGTTATG AAAATAAGGT	4560
35	GAAAGAAAAAT CTATTACAAC GTGCACAAAC CTACAATATG TTGGATAATA TTCTACGGCT	4620
	TGAAATTCCA ACACAAACAG TGCAAGTTGA AAAAAATGGA AAGAGAAAAA AAGTAGAAGA	4680
	AAATCGCTTT CCAGGTTATG TTCTTGTAAG AATGGTCATG ACAGATGAAG CTTGGTTTGT	4740
40	TGTTCGAAAC GCACAGAGTC CTACAAAATT CATTTCAGAA CAAACAGCTT ATGAAATTGA	4800
	TGAAGAGGTT CGTTCAATTAT TAAATGAGGC ACGAAATAAA GCTGCTGAAA TTATTGATC	4860
	AAATCGTGAA ACTCACAAGT TAATTGCAGA AGCATTATTG AAATACGAAA CATTGGATAG	4920
45	TACACAAATT AAAGCTCTTT ACGAAACAGG AAAGATGCCT GAAAGCAGTA GAAGGGAAT	4980
	CTCATGCACT ATCCTATGAT GAAGTAAAGT CAAAATGAA TGACGAAAAA TAACCTGAG	5040
50	AGAGGCTGGA GCCTCTCTTT TTTGTGAGT TTAGGAGCTA AAGGGAACAG AATGGAGAAA	5100
	ATGGAACAAA TGTGTTTCTT AATCTGTTAG ACTGTATCTA GAAAGGGGAA AATTATGATT	5160
55	AAGAATTGTG ATGAAGAAGT CCAAGGGAAT GTGTATAAGT GTAGAAATGA ATATTACCTT	5220
	CATTATGGG AATTGTGCGA TTGGGACCAA GAAGGCATGC TCTGCTTACA TGAATTGATT	5280
	AGTAGAGAAG AAGGACTGGT AGACGATATT CCACGTTTAA GGAAATATTT CAAAACCAAG	5340
60	TTTCGAAATC GAATTTTAGA CTATATCCGT AAGCAGGAAA GTCAGAGCG TAGATACGAT	5400



WO 98/26072

PCT/US97/22578

- 75 -

AAAGAACCCCT ATGAAGAAGT GGGTGAGATC CCCGGTACCG AGCTCGAA 5448

(2) INFORMATION FOR SEQ ID NO:30:

5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1040 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

15 (iv) ANTI-SENSE: NO

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TAGTGAATTC	GAGCTCGGTA	CCCGGGGATC	GTTTCTCGGT	TCTTTTGGAG	CACAAGGGCA	60
TCCATCCCAT	TGCTATATTT	TCCAAAATGG	ATTGTGTGGA	AGATAGGGGA	GAACGTGGATT	120
TTTACCAGCA	GACCTATGGT	GACATCGGCT	ATGACTTTGT	GACCACTAAA	GAGGAACCTCC	180
TGTCTTTGTT	AACAGGCAAG	GTTACGGTCT	TTATGGGGCA	GACAGGTGTT	GGGAAGTCAA	240
CTCTTCTCAA	TAAATTCGCA	CCAGACCTCA	ATCTTGAAAC	GGGAGAAATT	TCAGACAGTC	300
TAGTTCGCGG	TCGCCATACC	ACTCGAGCTG	TTAGTTTTTA	CAATCTCAAC	GGGGGTAAAA	360
TCGCAGATAC	ACCAGGATTT	TCATCCTTGG	ACTATGAAGT	ATCAAGGGCT	GAAGACCTCA	420
ATCAGGCTTT	CCCAGAGATT	GCTACTGTTA	GCCGAGATTG	TAACTTCCGT	ACTTGTACCC	480
ATACCCATGA	GCCGTCTTGT	GCCGTCAAAC	CAGCTGTTGA	AGAGGGTGTT	ATTGCAACCT	540
TCGGTTTTGA	CAATTACCTG	CAATTCCTTA	GTGAAATTGA	AAATCGTAGA	GAAACCTATA	600
AAAAAGTCAG	CAAAAAAATT	CCAAAAATAAG	GAGAAACCTA	TGTCTCAATA	CAAGATTGCT	660
CCGTCAATTC	TGGCAGCAGA	TTATGCCAAC	TTTGAACGTG	AAATCAAACG	TCTAGAAGCA	720
ACTGGGGCAG	AATATGCCCA	TATCGATTCT	GGCAGTCAT	TTTGTACCGC	AAATCAGTTT	780
TGGTSCAGGT	GTGGTCGAGA	GCTTCGTCT	CATAGTAAGA	TGTTTTTCGA	TTGCCACTTG	840
ATGGTGTCAA	ACCCTGAGCA	TCATCTGGAA	GATTTTGGCG	GTGCAGGTGC	AGACATCATC	900
AGTATCATATG	TAGAAGCAAC	ACCTCATATT	CATGCGCGCC	TCCAAAAAAT	TCGTTCACTC	960
GGAGTTAAGC	CTTCAGTCGT	TATCAATCCT	GGCACACCAG	TTGAAGCCAT	CAAGCACGTC	1020
CTTCATCTAG	TGACAAGTTT					1040

(2) INFORMATION FOR SEQ ID NO:31:

60 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 789 base pairs

WO 98/26072

PCT/US97/22578

-76-

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

5 (11) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(1v) ANTI-SENSE: NO

10

(xi) SEQUENCE DESCRIPTION: SEO ID NO:31:

15	ATATCACGAC	GGAGCCATAC	TACCGATTTT	CTTAGACATA	GCGCCACCTT	TACCGATGAT	60
	AATCCCTTTT	TGGGTATGCG	GCTCGACCAT	GATGTGTTGA	CSGATGTGAA	CCTGTCTGT	120
20	CTCTTCTGCT	CGTTTCATAG	AGTCAACAAC	TACTGCTACA	GAATGCGGAA	TCTCTTCACG	180
	AGTTAGGTGC	AAGACTTTCT	CGCGAACCAT	TTCTGAAACT	AAGAAACGTT	CTGGATGATC	240
	TGTGATTTGA	TCAGACGGGA	AATATTGGAA	ACCTTCATCC	AGATTTTCAC	TCAAAATATC	300
25	CACCTAGACG	GACACGTTAT	TTCCCTGAAG	GGCTGAGATT	GGAAACAATT	CCTTAAAGTC	360
	CATTTGATTA	CGGAAGTCAT	CAATCTGAGA	CAAGAGCTGG	TCTGGATGGA	CCTTATCGAT	420
30	TTTATTCAAC	ACCAAAATCA	CAGGAACCTT	GGCAGCCTTG	GAGACGCTCG	ATAATCATAT	480
	CGTCCCCTTT	ACCACGCGCT	TCATCAGCAG	GCACCATGAA	AAGAACAGTG	TCCACTTTCG	540
	GAAGGTACTG	TAGGCAAGACT	CAACCATGAA	ATCTCCGAGA	GCTGTTTTAA	GTGTGTGAAT	600
35	CCCTGGTGTG	TCGATAAAGA	CAATTTGCTC	CTTATCAGTC	GTGTAATTCC	CATGATTTTA	660
	TTGCGCGTTG	TCTGCGCCTT	GTCACTCATG	ATGGCAATCT	TTTGCCCCAT	AACGTGATTT	720
40	AAAAAGSTTG	ACTTCCCAAC	ATTGGGACTC	TACAAATGCG	TACAAACCTG	ATTTAAAAAT	780
	CATAATTCCT						789

(2) INFORMATION FOR SEQ ID NO:32:

45 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1233 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
50 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

60

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:32:

-77-

	TGAGATAAAC TTGCGACTCA TATGAGAATA TGAATCAAGC CGTCCTCGTG AACCCCGATA	60
5	TCAACGAAGG CACCGAAGTC AACCAACGTTA CGCACACAC CTCTAGCTT CTGACCTACC	120
	ACTAGGTCTT TGATATCTAG GACATCTTGG CGAAGCACAG GTGCGTCAAA GGAATCACGG	180
	AAATCTCGAC CTGGTTTGAG AAGATCTGCA ATGATATCTT TAAAGGTTTC TGGACCGAGG	240
10	TCTAGCTCTT GAGCGATTTT CTTGACTGAA AGGGACTTGA GTTTGCTTGG GCTTCTTCGT	300
	TTAGGTCCTT AATATCTAAA CGTTTGAAGA GTCCTTAACT GCAGTGTAAT TCTCTGGGTG	360
15	AACCTCTGTA TTATCAAGGA TATTGCTACT TTCAGGGATA CGAAGGAAC CAGCAGCCTG	420
	CTCAAAGGCC TTGGCTCCCA GACGAGGAAC TTTCTTGATT TGGCGCGGTG AAGTGATTTT	480
	TCCTTCTTCC TCGCGGTATT TGACAAATTT TTCAGAGATA GTTTTGTTGA GTCCAGCTAC	540
20	GTGTGAAGA AGAGCTGGGC TAGCTGTATT GACATTGACA CCAACTGGT TAACCACTGT	600
	ATCGACAACA AAGTCCAGAC TCTCAGATAG TTTCTTCTGA CTGACATCGT GTTGGTATTG	660
25	ACCGACACCA ATTGACTTAG GATCGATTTT GACCAATTCC GCAAGAGGAT CTGCAAAACG	720
	ACGGGCGATA GAAATGGCAG AGCGTTTTTC AACGGTCAAG TCTGGAATC CCTGACGAGC	780
	AAGTTCGCTG GCAGAAATAGA CAGAAGCACC ACTTTCATTA ACGATAACAT AGCTGACTTC	840
30	AGGGAAATCT TTCAGAACTT CCGCTACAAA AGCTTCACIT TCACGACTGG CCGTTCATT	900
	TCCAATGGCA ATAATCTCTA CACCGTATTG ACCAATTAAA TCTGCTAAAT CTTTCTTGGC	960
35	TTCTTCGATT TGACGAGCTG ATGCTGGTTT AACAGGATAA ATAACCTGAG TTGTACGAT	1020
	TTTTCTCTTT GCATCCACGA CAGCTAACTT GGCACCTGTA CGAAAGGCTG GGTCAAATCC	1080
	AAGAACCACG CGCCCTTTCA GTGGAGCAAC CAAGAGGAGA TTGCGCAGAT TGTCAAGAAA	1140
40	AAGTTGGATA GCTCCCTCTT CAGCTTTCTC AGTTAATTCT GTCCGAATAC GACGCTCAAT	1200
	AGCAGGCAAG ACCTTTTTTCT TAACGGATTG CTG	1233

## (2) INFORMATION FOR SEQ ID NO:33:

- 45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6679 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear
- 50 (ii) MOLECULE TYPE: DNA (genomic)
- 55 (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- 60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

-78-

	ACAAAGGCGTC ATCCGCTGAT TTCTTAAATA CGGACCAAAT GGTGAGAAAG TTATCACTAA	60
5	CTTGAAACGT GTTTCATAAC CAGGACTTCG TGCTACAAA AAACGTGAAG ACCTTCCAAA	120
	AGTTCCTAAC GGACTTGGAA TTGCCATCCT TTCAACTTCT GAAGGTTTGC TTA CTGATAA	180
	AGAAGCACGC CAAAAGAATG TTGGTGGTGA GGTATCGCT TACGTTTGGT AAAATCAAGA	240
10	TACAAAGCTC GTAAAGAACA AAGCAAAATT AGGAAGTTGG AGAAGTTTGT TTACAAACAA	300
	GCCAACTTAT CTATTTTGCA CAGTCTCTAG ATCGTGTTCA GTTCAGCTCT TGAACATAAT	360
15	AAGTATCTGA ACCCCGTGAA AACTGGCCGT TCTGGCTGAC AATTTAACAG GAGAAAATAA	420
	ACATGTCACG TATTGGTAAT AAAGTTATCG TGTTGCCTGC TGSTGTTGAA CTCGCTAACA	480
	ATGACAAAGT TGTAACTGTA AAAGGACCTA AAGGAGAACT TACTCGTGAG TTCTCAAAAG	540
20	ATATTGAAAT CCGTGTGAA GGTACTGAAG TAACTCTTCA CCGTCCAAAC GATTCAAAAG	600
	AAATGAAAC TATCCACGGA ACTACTCGTG CCGTTTGA CAACATGGTT GTTGGTGAT	660
25	CAGAAGGATT CAAGAAAGAA CTTGAAATGC GTGGGGTTGG TTACCGTGCA CAGCTTCAAG	720
	GATCTAAACT TGTTTTGGCT GTTGGTAAAT CTCATCCAGA CGAAGTTGAA GCTCCAGAAG	780
	GAATTACTTT TGAACCTCCA AACCCAACAA CAATCGTTGT TAGCGGAATT TCAAAAGGAA	840
30	TAGTTGGTCA AACAGCTGCT TACGTACGTA GCCTTCGTTT ACCAGAACCA TATAAAGGTA	900
	AAGGTATCCG TTACGTTGGT GAATTCGTTG TCGTAAGA AGGTAAACA GGTAAATAAT	960
35	GTGAGTGGT TGATCATCAA CCACCAACCT ATTTTCCAAC TTTGTGCATA GCAACGATTT	1020
	AAACTAAAG AGGTGAAAC TGATGATTTCA AAACAGATA AAAACAACT CCGCCAAAA	1080
40	CGCCACCGTC GCGTCCGGA AAATCTCTCG GAATCGTGA TCGCCACGT TTGAACGTAT	1140
	TCCGTTCTAA TACAGGCATC TACGCTCAAG TGATTGATGA CGTAGCGGGT GTAACGCTCG	1200
	CAAGTGCTTC AACTCTTGAT AAAGAAGTTT CAAAAGGAAC TAAAACTGAA CAAGCCGTTG	1260
45	CTGTCGGTAA ACTCGTTGCA GAACGTGCAA ACGCTAAGG TATTTAGAA GTGGTGTTCG	1320
	ACCGCGGTGG ATATCTATAT CACGGACGTG TGAAGCTTTT GGCTGATGCA GCTCGTAAA	1380
	ACGGATTGAA ATTCTAATAG GAGGACACTA GAAATGGCA TTTAAAGACA ATGCAAGTTGA	1440
50	ATTAGAGGAA CGCGTAGTTG CTGTCAACCG TGTTACAAA GTTGTAAAG GTGGACGTCG	1500
	TCCTCGTTTC GCAGCTCTTG TTGTTGTTGG TGACCAAAAT GGTGCGGTAG GATTGGTATC	1560
55	TGTTAAAGCT CAAGAAGTTC CAGAAGCAAT CCGTAAAGCA GTAGATGATG CTAAGAAAAA	1620
	CTTGATCGAA GTTCCTATGG TTGGAACAAC AATCCCACAC GAAGTCTTTT CAGAATTCTG	1680
	TGGAGCTAAA GTATTGTTGA AACCTGCTGT AGAAGGTTCT GGAGTTGCCG CTGGTGGTGC	1740
60	AGTTCGTGCC GTTGTGGAAT TGCGAGSTGT GGCAGATATT ACATCTAAAT CACTTGGTTT	1800



	TGTTGATTAT TCTCTTTACA TTCTTCTATA CGTTTGTACA GATTAATCCT GAAAAAGCAG	3720
5	CACAGAGCCT ACAAAGAGT GGTGCCTATA TCCATGGAGT TCGTCTCGT AAGGTACAG	3780
	AAGAATATAT GTCTAAACTT CTTGTCGTG TCGCACTGT TGGTCCCTC TTCTTGGTG	3840
	TGATTCCCAT TTTACCGATT GCAGCTAAAG ATGTATTGG TCTTCTGAT GTTGTTCGCT	3900
10	TTGTGGAAAC AAGTCTCTTG ATCATTATCT CTACAGGTAT CGAAGGAATC AAGCAATTGG	3960
	AAGGTACCT ATTGAAACGT AAGTATGTTG GTTTCATGGA CAGAACAGAA TAAAGTATT	4020
15	TACTGAATCA GTAAATACTG AGGGAGTGGG GGTTAACACT CTGACATTGG TAAGAGTTGG	4080
	ATCTCCCCTC TTCTATTTTG TTTTAAATC GGGGTGAAAA AACTTTTTCG TTCTATTAA	4140
	AAACAAAATA AGGAGATCAA ATCATGAATC TTTTGATTAT GGGCTTACCT GGTGCAGGTA	4200
20	AGGGAACTCA AGCAGCAAAA ATCGTAGAAC AATTCCATGT TGCACATATC TCAACAGGTG	4260
	ATATGTTCCG CGCTGCAATG GCAAAATCAA CTGAAATGGG TGTCTTGCT AAGTCATATA	4320
	TTGACAAGGG TGAATTGGTT CCTGACGAAG TTACAAATGG AATCGTAAAA GAACGCCCTT	4380
25	CACAGATGA TATTAAAGAA ACAGGATTCT TATTGGATGG TTACCCACGT ACAATTGAAC	4440
	AAGCTCATGC CTTGGACAAA ACATTGGCTG AACTTGGCAT TGAACTAGAA GGTATTATCA	4500
30	ATATTGAAGT GAACCCCTGAC AGCCTCTTGG AACGTTTGAG TGCCCGTATC ATCCACCGCG	4560
	TAACTGGAGA AACTTTCAC AAGGTCTTTA ACCCACCAGT TGACTATAAA GAAGAAGATT	4620
35	ACTACCAACG TGAAGATGAT AAGCCTGAGA CAGTAAAAAG TCGTTTGAT GTTAATATTG	4680
	CTCAAGGAGA ACCAATCATT GCTCACTACC GTGCCAAAGG TTTGTTTCAT GACATCGAAG	4740
	GTAAATCAAGA TATCAATGAT GTCTTCTCAG ATATTGAAAA AGTATTGACA AATTTGAAAT	4800
40	AAAGCGTTTT TCACACTTGC AAAAATCCGC TACAAATGTT ATACTGAAAT AGTCTGACTT	4860
	ATAATGTTG TCTCTGTGT TAGAGGCATC GAATCGAAAT TTATGGAGGT GCTTTTSGGT	4920
45	GGCAAAAGAC GATGTGATTG AAGTTGAAGG CAAAGTAGTT GATACAATGC CGAATGCAAT	4980
	GTTTACGGTT GAACCTGAAA ATGACATCA GATTTTAGCA ACAGTTTCTG GTAAAAATTCG	5040
	TAAAACTAT ATTCTGATTT TAGCGGGAGA TCGTGTTACT GTCGAAATGA TGCCATATGA	5100
50	CTTGACACGT GGACGTATCA CTTACCGCTT TAAATAATCG AAAAATCTGG AGGGATAAGA	5160
	AATGAAAGTA AGACCATCGG TCAAAACCAAT TTGCGAATAC TGTAAAGTTA TTCGTCTGTA	5220
55	TGGTCTGTAT ATGATAATTT GCCCAGCAAA TCCAAAACAC AAACAACGTC AAGGATAAGA	5280
	TAGAAAAGAG AAAACATGGC TCGTATTGCT GGAGTTGATA TTCCAAATGA CAAACGCGTA	5340
	GTAAATCTCAT TGACTTATGT TTATGATATC GGACTTGCAA CATCTAAGAA AATTTTGGCT	5400
60	GCTGCTGGAA TCTCAGAAGA TGTTCTGTGA CGTGATCTTA CATCATGATC AGAAGATGCT	5460

WO 98/26072

PCT/US97/22578

- 81 -

```

ATCCGTCGTG AAGTGGATGC AATCAAAGTT GAAGGTGACC TTCGTCGTGA AGTAAACTTG 5520
AACATCAAAC GTTTCATGGA AATCGGTTCA TACCGTGGA TCCGTCACCG TCGTGGACTT 5580
5 CTTGTCCTGGT GACAAAAATAC TAAAAACAAC GCTCGCACTC GTAAAGGTAA AGCTGTTGCG 5640
ATTGCTGGTA AGAAAAATA ATATAGGAGG TAAAGTCTT GGCTAAACCA ACACGTAAAC 5700
GTCGTGTGAA AAGGAATATC GAATCTGGTA TTGCTCATAT TCACGTAGCA TTTAATAACA 5760
10 CTATTGTTAT GATTACTGAT GTGCATGGTA ATGCAATTGC TTGCTCATCA GCTGCTGCTC 5820
TTGTTTCAA AGGTTCTCGT AAATCTACAC CATTGCTGTC TCAAAAGGCT TCTGAAGCTG 5880
15 CTGCTAAATC TGCACAAGAA CACGGTCTTA AATCAGTTGA AGTTACTGTA AAAGGTCCAG 5940
GTTCTGGTCG TGAGTCAGCT ATTCGTGCGC TTGCTGCCGC TGGTCTTGAA GTAAACAGCA 6000
TTGCTGATGT GACTCCAGTG CCACACAATG GTGCTCGTCC TCCAAAACGT CGCGTGTAT 6060
20 AATCATCGCA TTACACTGCT TTTGCTTTAA GAGGGAGTAA CTAAATGATC GAGTTTGAAA 6120
AACCATAAT AAAAAAATAT GATGAAAAA AAGATTATGG CAAGTTTGTG ATCGAACACC 6180
25 TTGAACGTGG CTACGGTACA ACTCTTGGTA ACTCTCTTCG TCGTGTACTT CTAGCTTCTC 6240
TACCAGGAGC AGCTGTGACA TCTATCAACA TTGATGGTGT GTTACATGAG TTTGACACAG 6300
TTCCAGGTGT TCGTGAAGAC GTGATGCAAA TCATTCTGAA CATTAAAGGA ATTGCAGTGA 6360
30 AATCGTACGT TGAAGACGAA AAAATCATCG AACTGGATGT TGAAGGTCC TCTGAAGTAA 6420
CAGCTGGTGA CATTTTGACA GATAGCGATA TTGAAATTGT AAATCCAGAT CATTATCTCT 6480
35 TTACAATCGG TGAAGGTTCT TCTCTAAAAG CGACTATGAC TGTTAACAGT GGTCTGGAT 6540
ATGIACCTGC TGACGAAAAA AAAAAGGATA ATGCACCAGT TGAACACCTT GCTGTAGATT 6600
CTATTATAC ACCAGTTACA AAAGTCAACT ATCAAGTGA ACCTGCTCGT GTAGGTAGCA 6660
40 ATGATGGTTT CGACTCTAG 6679

```

(2) INFORMATION FOR SEQ ID NO:34:

```

45 (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 1703 base pairs
      (B) TYPE: nucleic acid
      (C) STRANDEDNESS: single
      (D) TOPOLOGY: linear
50 (ii) MOLECULE TYPE: DNA (genomic)
      (iii) HYPOTHETICAL: NO
55 (iv) ANTI-SENSE: NO
60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```

WO 98/26072

PCT/US97/22578

- 82 -

	AGAATACCTT GGGGCAACTG TTCAGTCAT TCCTCATATC ACAGATGCTT TGAAGAAAA	60
	AATCAAGAGT GCCCGCTTAA CGACCGACTC TGATGTCATT ATCACAGAGG TTGGTGAAC	120
5	AGTAGGAGAT ATCGAGTCTT TGCCATTCTT AGAGGCTCTT CGCAGATGAA GGCAGATGTG	180
	GGGCGGATAA TGTCATGTAT ATCCATACAA CCTTCTTCTT TACCTCAAGG CTGCTGGTGA	240
10	AATGAAACCA AACCAACCA ACACCTCTGT AAAAGATTGGG TGGCTTGGGA ATCCAACCAA	300
	ATATGTTGGT TATTCGTACA GAAGAGCCAG CTGGTCAAGG AATTAAAAAT AAATGCCCC	360
	AGTTCTGTGA TGTGGCACCA GAATCCCTAA TCGAATCGTT GGATGTTGAA CACCTTTACC	420
15	AAATCCACT GAACTTGCAG GCACAAGGGA TGGACCAAT TGTTTGTGAT CATTTGAAAT	480
	TAGACGCACC AGCAGCGGAT ATGACAGAAT GGTACGCCAT GGTGGACAAG GTCATGAACC	540
20	TCAAGAAACA AGTTAAGATT TCCCTTGTG GTAAGTATGT GGAGTTGCAA GATGCCATATA	600
	TCTCAGTGGT CGAAGCCTTG AAACACTCTG GCTATGTCAA TGATGTAGAA GTTAAATCA	660
	ATTGGGTCAA TGCCAATGAT GTGACAGCAG AGAATGTAGC AGAATCTTGT TCTGATGCGG	720
25	ACGGGATCAT CGTACAGGTT GGTTTTGGTC AACGTGGTAC AGAAGGGAAT ATCCAAGCCA	780
	TCCGCTATGC GCGTGAATAT GATGTTCCAA TGTGGGAGT CTGCTTGGGA ATGCAAGTTG	840
30	CATGATCGA GTTTCGCTGT CACGTTTATG GTCTTGAAGG TGCCAATTCT GCAGAGCTTG	900
	CACCAGAAAC AAAATACCTT ATCATTTGATA TCATGCGTGA TCAGATTGAT ATTGAGGATA	960
	TGGGTGAAC CCTTCGTTTG GGACTTTATC CGTCTAAGTT GAAACGTGGC TCTAAGGCTG	1020
35	CTGCTGCTTA TCACAATCAA GAAGTGGTGC AACGCCGTCA CCGTCACCGT TATGAGTTTA	1080
	AATAATGCCT TCCGTGAGCA GTTTGAGGCA GCAGGTTTGT CTTTTCAGGA GTTTCTCCAG	1140
40	ACAATCGTTT GGTGAAATC GTGGAATCC TGAATAATAA TTCTTTGTAG CTGTGCACTA	1200
	TCACCCTGAA CTGTACGGC TCCAACCGAC CAGAAGAACT CTACACTGCC TTTGTTACTG	1260
	CAGCGGTTGA GAACAGCAAT TAGCAAAATC AGAACCTTGT AGAAAAATCT CAGAGGTTTT	1320
45	TTGCATACGA TGATATTGCA GTATATCTGA GGTAGGAGTC CTCTGTATGT ACCTGCTACC	1380
	GTGAAATCA ATAGCGACTC CCTCTTGCCC TGTGCTAGTG AATGGATTTA TCAGTATATT	1440
50	GAAATGAAAT AAAATTTGAA CAAATTAATT CGGAAAGCCA AATCAATTTT TAGCAAGTT	1500
	TTAGGAACTG GATTGTATAG TGAATTGAAA TAAGATGTGA ACATCTCTAT CAGGAAAGTC	1560
	AAATTAATTT ATAGAAATAT TTTAGCAGTC AAGATGGACT GTTATAGATT CAATATACTA	1620
55	TACTTTTTTA ATTAAATCCA CTATAATAAA ATGAAATAAT AACAGGACAA ATCGTTACGG	1680
	ACAGTCAAAAT CGACTCTAGA GGA	1703

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:



WO 98/26072

PCT/US97/22578

- 63 -

(A) LENGTH: 1620 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

10

(iv) ANTI-SENSE: NO

15

(xi) SEQUENCE DESCRIPTION: SEQ ID No:35:

18	ATTGTAAAC ACCAAGGAAA AACAGCTAAA GAAGCGAAAG AATTGGCCAT TGACTACATG	60
20	AATAAGGTTG GCATTCCAGA CGCAGATAGA CGTTTAAATG AATACCCATT CCAATATTCT	120
	GGAGGAATGC GTCAACGTAT CGTTATTGCG ATTGCCCTTG CCGCCGACC TGATGTCTTG	180
	ATCTGTGATG AGCCAAACAAC TGCCTTGGAT GTAACATATC AAGCTCAGAT TATTGATTTG	240
25	CTAAAATCTT TACAAAACGA GTATCATTTT ACAACAATCT TTATTACCCA CGACCTTGGT	300
	GTGGTGGCAA GTATTGCGGA TAAGTAGCGG GTTATGTATG CAGGAGAAAT CGTTGAGTAT	360
30	GGACGGTTG AGGAAGTCTT CTATGACCCG CGCCATCCAT ATACATGGAG TCTCTGTCT	420
	AGCTTGCCCTC AGCTTGCTGA TGATAAAGGG GATCTTTACT CAATCCAGG AACACCTCCG	480
	TCACCTTATA CTGACCTGAA AGGGGATGCT TTTGCCCTTG GTTCTGACTA CGCAATGCAG	540
35	ATTGACTTCG AACAAAAAGC TCCTCAATTC TCAGTATCAG AGACACATGG GGCTAAAACT	600
	TGGCTCTTTC ATGAGGATGC TCCAAAAGTA GAAAAACCGA CTGTGATTGC AATCTCCAT	660
40	GATAAGATCC GTGAAAAAAT GGGATTGGCC CATCTGGCTG ACTAGGAGGA AGGAAATGTC	720
	TGAAAAATTA GTAGAAATCA AAGATTAGA AATTTCCTTC GGTGAAGGAA GTAGAAAGTT	780
	TGTCGGGGTT AAAAATGCTA ACTTCTTTAT CAACAAGGGA GAAACTTTCT CGCTTGATAG	840
45	TGAGTCCGGT AGTGGGAAAA CAACTATTGG TCGTGTATC ATCGGTCTAA ATGATACAAG	900
	TAATGGAGAT ATCAATTTTG ATGGTCAAAA GATTAATGGT AAGAAATCGC GTGAACAAGC	960
50	TGCGGAATTG ATTCTCGGAA TCCAGATGAT TTTCGAAGC CCGCCGCAA GTTTGAATGA	1020
	ACGTGCGACT GTTGATTATA TTATTCTGTA AGGTCTTTAC AATCACCCTT TATTTAAGSA	1080
	TGAAGAAGAA CGTAAGAGAA AAGTTCAGAG TATTATCCGT GAAGTAGGTC TTCTTGCTGA	1140
55	GCACCTGACT CGTTACCCCT ATGAATTCTC AGCGCGTCAA CGTCAACGTA TCGGTATTGC	1200
	CGTGCCCTTG GTCATGCAAC CAGACTTTGT TATTGCAGAT GAGCCAATTT CAGCCITGGA	1260
60	CGTTTCTGTA CGTGCCCAAG TCTTGAACTT GCTCAAAAAA TTCCAAAAAG AGCTCGGCTT	1320
	GACCTATCTC TTATCGCCCC ATGACTTGTG GGTGTGTCGC TTTATTTCAG ATCGTATCGC	1380

WO 98/26072

PCT/US97/22578

- 84 -

AGTTATTAC AAGGTCGTTA TTGTAGAGGT TGCAGAAACA GAAGAATTGT TTAACAATCC 1440  
AATTCACCCA TATACTCAAG CCTTGCTTTC AGCGGTACCA ATCCCAGATC CAATCTTGA 1500  
5 ACGTAAGAAG GTCTTGAAGG TTTACGACCC AAGTCAACAC GACTATGAGA CTGATAAGCC 1560  
ATCTATGGTA GAAATCCGTC CAGGTCACCTA TGTTGGGCG AACCAAGCCG AATTAGCACG 1620

10 (2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:  
15 (A) LENGTH: 984 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:  
30 GTACCCGGGG ATCAGGTTTT ACGGATTCTT GAAGTTCTCT GTGGGCAGGA CCTCTGCA 60  
GTAAGAGTAA GAGTGATTCT ACAAGATTTA CTAGAAGCTA GAAAAATGTG GCAAGCTAAT 120  
GTCAGCTTTC AAAATGCCAT GGAATATCTG GTCTTGAAAG AAATATAAAC TCAAAAATGA 180  
35 ATGATAAAGA AAGGAAAGGG CTGTTTTATG GACAAAAAG AATTATTTGA CGCGCTGGAT 240  
GATTTTCCCT AACGATTATT GGTAACCTTG GCCGATGTGG AAGCCATCAA GAAAAATCTC 300  
40 AAGAGCCTGG TAGAGGAAAA TACAGCTCTT CGTTTGAAAA ATTCTAAGTT GCGAGAAGCG 360  
TTGGGTGAGG TGAAGCAGA TGCTCCTGTC AAGGCCAAGC ATGTTCTGTA AAGTGTCCGT 420  
CGCATTTACC GTGATGGATT TCACGTATGT AATGATTTTT ATGGACAACG TCGAGAGCAG 480  
45 GACGAGGAAT GTATGTTTTG TGACGAGTTG CTATACAGGG AGTAGGCATG CAGATTCAAA 540  
AAAGTTTTAA GGGGCAGTCT CCCTATGGCA AGCTGTATCT AGTGGCAACG CCGATTGGCA 600  
50 ATCTAGATGA TATGACCTT CGAGCTATCC AGACCTTGAA AGAAGTAGAT TGGATTGCTG 660  
CTGAGGATAC GCGCAATACA GGTCTTTTGC TCAAGCATT TGCATTTCC ACCAAGCAGA 720  
TCAGTTTTCA TGAGCACAA TGCCAAGGAA AAATTCCTGA TTTGATTGGT TTCTTGAAAG 780  
55 CAGGGCAAAG TATGCTCAG GTCTCTGATG CCGGTTTGCC TAGCATTTC GACCTGTGTC 840  
ATGATTAGT TAAGGCAGCT ATTGAGGAAG AAATTGCAAT TGTGACAGT CCAGGTGCCT 900  
60 CTGCAGGAAT TTCTGCCTTG ATTGCCAGTG GTTTAGCGCC ACAGCCACAT ATCITTTAGG 960

WO 98/26072

PCT/US97/22578

-85-

GTGTTTACC GAGAAATCA GGTC 984

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1554 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CTAGAGTCGA	AAAGACAAGC	AGGAGCGTAT	TTCCAAAGAA	ACCATGGAAA	TCTATGCCCC	60
GCTTGCCCAT	CGTTTGGGGA	TTTCCAGTGT	CAAAATGGAA	TTAGAAGACT	TGCTTTCCG	120
TTATCTCAAT	CCAACGGAGT	TTTACAAGAT	TACCCATATG	ATGAAGGAAA	AGCGCAGGGA	180
GCGTGAGGCC	TTGCTGGATG	AGGTAGTCAC	AAAATTAGAG	GAGTATACGA	CAGAACGTCA	240
CTTGAAAGGG	AAGATTTATG	GTCTGCCCAA	GCATATTTAC	TCAATTTTCC	GCAAAATGCA	300
GGACAAGAGA	AAACGGTTTG	AGGAAATCTA	TGATCTGATT	GCTATTCTGT	GTATTTTAGA	360
TACCCAAAGT	GATGTTTATG	CGATGCTTGG	TTACGTGCAT	GAATTTTGGA	AACCGATGCC	420
AGGTCGCTTC	AAAGACTATA	TCGCCAACCG	CAAGGCCAAT	GGTTATCAGT	CTATCCATAC	480
GACTGTTTAT	GGACCAAAAG	GGCCGATTGA	ATTCCAGATT	CGAACCAAGG	AAATGCACGA	540
GGTGGCTGAG	TACGGGGTTG	CGGCTCACTG	GGCTTATAAG	AAAGGTATAA	AGGGGCAAGT	600
TAACAGCAAG	GAATCAGCTA	TTGGAATGAA	CTGGATCAAG	GAGATGATGG	AGCTCCAAGA	660
CCAGGCTGAT	GATGCTAAGG	AAATTGTGGA	CTCTGTTAAG	GAAACTATT	TGGCTGAGGA	720
GATTACCGTT	TTACCCGAGA	TGGAGCTGTC	CGTTCCTTCC	CAAGATTCA	GGACCGATTG	780
ATTTTGCTTA	CGAAATCCAT	ACCAAGGTCT	GTGAAAAGCA	ACTGGTGCCA	AGGTCAATGG	840
CCGATGAGTT	CCACTGACAC	CCAAAGTTAA	GGACAGGGGA	TCAGGTTGAA	ATTATCGCCA	900
ACCCGAARCT	CTTTGGACCT	TAGCCGTGAC	TGGCTCAATA	TGGTCAAGAC	TAGCAAGGCG	960
CGCAATAAGA	TTGCGCAGTT	CTTTAAAAAC	CAAGATAAGG	AATTGTCTGT	CAACAAGGGT	1020
CGTGAGATGC	TGATGGCTCA	GTTCCAAGAA	AATGGCTATG	TGGCAATATA	ATTATGAGAC	1080
AAGCGCCACA	TGATCAAGT	TCTGCAAAAG	ACCAAGTTACA	AGACAGAAGA	CTCCCTCTTT	1140
GCGGCCATTG	GTTTGGGGA	AATCGGTGCG	ATTACCGTCT	TTAACCGTCT	GACTGAAAAG	1200

WO 98/26072

PCT/US97/22578

-86-

	GAGCGCCGTG AGGAAGAGCG TGCCAAGGCC AAGGCTGAGG CAGAGGAGCT TGTCAAAGGT	1260
	GGCGAGGTCA AGGTTGAAAA TAAAAAAACT CTCAGAGTCA AGCATGAGGG GGGAGTGGTT	1320
5	ATTGAAGGTG CTTCTGGTCT CCTAGTGC GG ATTGCTAAGT GTTGTAAACC CGTGCTGGT	1380
	GACGATATTG TTGGCTACAT TACCAAGGGT CGTGGTGGG CTATTCACCG TGTGGACTGT	1440
	ATGAACCTGC GTGCCAAGA AAACCTACGAG CAACGCTCC TTGATGTGGA ATGGGAAGAC	1500
10	CAGTACTCTA GCTCAATAA GGAGTATATG GCCCATATCG ACTCTAGAGG ATCC	1554

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3190 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

	CCTCAGTGCT AAAACAGGGG AAATCTGGC AACACGCAA CGACCGACCT TTGATGCAGA	60
	TACAAAAGAA GGCATTACAG AGGACTTTTT TGCGGTGATA TCCTTTACCA AAGTAACATAT	120
35	GAGCCAGGTT CCACTATGAA AGTGATGATG TTGGCTGCTG CTATTGATAA TAATACCTTT	180
	CCAGGAGGAG AAGTCTTTAA TAGTAGTGAG TTA AAAAATTG CAGATGCCAC GATTTCGAGAT	240
40	TGGGACGTTA ATGAAGGATT GACTGGTGGC AGAATGATGT CTTTTCTCA AGGTTTTGCA	300
	CACCTCAAGTA ACGTTGGGAT GACCTCCTT GAGCAAAAGA TGGGAGATGC TACCTGGCTT	360
	GATTATCTTA ATCGTTTTAA ATTTGGTGTT CCGACCCGTT TCGGTTTGAC GAGTAGATAT	420
45	GCTGGTCAGC TTCTCGCGGA TAATATTGTC AACATTGCGC AAAGCTCATT TGGACAAGGG	480
	ATTTCACTGA CCCAGACGCA AATGATTGCT GCCCTTACAG CTATTGCTAA TGACGGTGTC	540
50	ATGCTGGAGC CTAAATTTAT TAGTGCCATT TATGATCCAA ATGATCAAAAC TGCTCGGAAA	600
	TCTCAAAAAG AAATTGTGGG AAATCCTGTT TCTAAAGATG CAGCTAGTCT AACTCGGACT	660
	AACATGGTTT TGGTAGGGAC GGAATCCGGTT TATGGAACCA TGTATAACCA CAGCACAGGC	720
55	AAGCCAACTG TAACTGTTCC TGGGCAAAAT GTAGCCCTCA AGTCTGGTAC GGCTCAGATT	780
	GCTGACGAGA AAAATGCTGG TTATCTAGTC GGGTTAACCG ACTATATTTT CTCGGCTGTT	840
60	CGATGAGTCC GGCTGAAAT CCTGGATTTT ATCTGTATG TGACGGTCCA ACAACCTGGA	900

WO 98/26072

PCT/US97/22578

-87-

	ACATTATTCA GGTATTCACT TGGGAGAATT TGCCAATCCT ATCTTGAGC GGGCTTCAGC	960
	TATGAAAGAC TCTCTCAATC TTCAAACAAC AGCTAAGGCT TTGGAGCAAG TAAGTCAACA	1020
5	AAGTCCCTTAT CCTATGCCCA GTGTCAAGGA TATTTCACCT GGTGATTAG CAGAAGAATT	1080
	GGTTCGCAAT CTGTGTACAA CCATCGTTGT GGGAACAGGA ACGAAGATTA AAAACAGTTC	1140
10	TGCTGAAGAA GGGAGAATC TTGCCCGAA CCAGCAAGTC CTTATCTTAT CTGATAAAGC	1200
	AGAGGAGGTT CCAGATATGT ATGGTTGGAC AAAGGAGACT GCTGAGACCC TTGCTAAGTG	1260
	GCTCAATATA GAACTTGAAT TTCAAGGCTC GGGCTCTACT GTGCAGAAGC AAGATGTCG	1320
15	TGCTAACACA GCTATCAAGG ACATTAAAAA AATTACATTA ACTTTAGGAG ACTAATATGT	1380
	TTATTCCCAT CAGTGTGGG ATTGTGACAT TTTTACTAAC TTTAGTAGGA ATTCGGGCT	1440
20	TTATCCAATT TTATAGAAAG GCGCAAAATTA CAGGCCAGCA GATGCATGAG GATGTCAAAC	1500
	AGCATCAGGC AAAAGCTGGG ATTCCTACAA TGGGAGGTTT GGTTTTCTTG ATTACTTCTG	1560
	TTTTGGTTGC TTTCTTTTC GCCCTATTTA GTAGCCAATT CAGCAATAAT GTGGGAATGA	1620
25	TTTTGTTCAT CTTGGTCTTG TATGGCTTGG TCGGATTTTT AGATGACTTT CTCAGGCTCT	1680
	TTCTGAAGAT CAATGAGGGG CTTAATCCTA AGCAAAAATT AGCTCTTCAG CTTCTAGGTC	1740
30	GAGTTATCTT CTATCTTTTC TATGAGCGCG GTGGCGATAT CCTGTCTGTC TTTGGTTATC	1800
	CAGTTCAATT GGGATTTTTT TATATTTTCT TCGCTCTTTT CTGGCTAGTC GGTTTTTTCA	1860
	ACGCAGTAAA CTGACAGAC GGTGTTGTAC GGTTTAGCTA GTATTTCCGT TGTGATTAGT	1920
35	TTGTTGCGCT ATGGAGTTAT TGCCTATGTG CAAGGTCAGA TGGATATTCT TCTAGTGATT	1980
	CTTGCCATGA TTGGTGGTTT GCTCGGTTTC TTCATCTTTA ACCATAAGCC TGCCAAGGTC	2040
40	TTTATGGGTG ATGTGGGAAG TTTGGCCCTA GGTGGGATGC TGGCAGCTAT CTCTATGGCT	2100
	CTCCACCAGG AATGGAATCT CTGATTATC GGAATTGTGT ATGTTTTTGA AACAACTTCT	2160
	GTATGATGC AAGTCAGTTA TTCAAACCTG ACAGGTGGTA AACGTATTTT CCGTATGACG	2220
45	CCTGTACATC ACCATTTTGA GCTTGGGGGA TTGTCTGGTA AAGGAAATCC TTGAGGCGAG	2280
	TGGAAGGTTG ACTTCTCTTT TTGGGGAGTT GGGCTTCTAG CAAGTCTCCT GACCCTCGCA	2340
50	ATTTTGTATT TGATGTAAGA ATGGCACCCT GATGTTTCAG GGTGTTTTTG TGTTTAAATA	2400
	CACAATGAAA ATCAAGAAGC AACTAGAAAA GCTAACTTTA GGCTGCTCAA AATATAATAT	2460
	ATTGAACTA GAATAGTACA CCTCTACTTC TAAACATTG TTGAAATCG ATTTGACTGT	2520
55	CCTGAACGAT TTATCCTGTT CTTATTTTAT TTTACTATAC AGTTTCGAGG TTGTAGATAA	2580
	GGCGAAGCTG ATGTGGTTTG AAGAGATTTT CTGAAAAGTG TTAACACCTA CAGACAAGCC	2640
60	TGACGATAGC AAGAAGTACC CTACTCGATA GGTATCGGCT TTTGCTTTCT GAAAAAATT	2700
	ATTTTAAAGCA TTTGACAAAT CTAGCAACAA AAAATCTAT AAATATAATA GATTGAAACT	2760

WO 98/26072

PCT/US97/22578

- 88 -

	AGAATAGTAC ACATCTACTT CTA AACATT GTTAGAAATC GATTTGACTG TCCTGATCGA	2820
5	TTTGCTCTGT TCTGTTCATA TTTTACTATA TTTCTATGAT AAAACGCATA GTATCAAGTT	2880
	TTCTTAATCC CCGATACTA TCGGTGTTG TAATTTTAA GATTTTGTGC TTAGAGTCGA	2940
	CTCCTTATTT TAGATATTTA AAAGGAATCT CACTTCCACA GAGCCAGTTG TAGACTTGGT	3000
10	CATTAAACAA TACATTTCATG GCTTCGTGAG CATACTCAGG CATGATACGA TAGGTTTTAT	3060
	CGCAGGTCAG ACGATTATAA ATCGCAAACT GGGTAATGGG ATAGCAAAACA TCGTCGTCCA	3120
15	AGCCCGTAAT CATCTTAACC TCACCTTGGA TACGATGGGC AAGATTTTTG ACATCGACTC	3180
	TAGAGGATCC	3190

(2) INFORMATION FOR SEQ ID NO:39:

20	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 5992 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: single
25	(D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: DNA (genomic)
	(iii) HYPOTHETICAL: NO
30	(iv) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:39:

35	TTGTCTTAG TGTTCCGACA AAGATTCTTC AAAATCAAT CATGGAAGAA GAAGGTAAC	60
40	GTCTCAAGGA AGTGTTCCAT ACAGATATTC ATAGCTTAAA GGGACCACAA AATTATCTGA	120
	AGTTGGATGC CTTTATCAT TCTTGCAAGA AAATGATGAA AATCGCTTAT TTAGACGCTT	180
	TAAATGCAA GTCTTGCTCT GGCTTACTGA GACAGAGACA GGAGATTGG ATGAAATCGG	240
45	GCAACTCTAC CGTTACCAAC ATTTTCTAGC AGACCTTCGT CATAATGGGA ATTTATCATC	300
	CCAGAGCTTA TTTGTGACGG AAGATTTTTG GAAACGTAGT CAAGAAAGGG CAGAGACTTG	360
50	CAAGCTTTTA GTGACTAATC ATGCCTATCT CGTAACCAGA CTTGAAGATA ATCTGTAATT	420
	TGTCAGTGAC CGTTTACTGA TTATTGATGA AGTCCAAAAG ATTTTGTTAG CTCTAGAAAA	480
	TCTGCTTCAA GAGACCTACG ATATACAATC TATTATCGAT TTAATTGATA AGGCTTTAGT	540
55	AGSAGAAGAA AACAGGGTTC AACACGGAT ACTAGAAAGT ATTCGCTTTG AGTGCCTCTA	600
	CTTGATAGAA CAATTTCAGT CTGGCAATC TAGGAAAAAT ATCTTAGATT CTCTGGACAA	660
60	TCTCCATCAG TATTTTTTCAG AATTAGAAGT GGAAGGCTTT GATGAGCTGG TTCGTATTTT	720
	TACAGCTGAA GGTGATTACT GGCTTGAAGT AACTGAAACG AGTCAAAAGA AAATTACAGAT	780

WO 98/26072

PCT/US97/22578

- 89 -

	TTCTTCTACA	AAATCAGGCC	GFACTCTTCT	GTCTCTTTA	CTTCTGAGA	GTTGCCAAGT	840
5	CTTGGGAGTA	TCGGCTACTC	TTGAGATTAG	TCAGAGGGTT	TCTTTGGCAG	ACCTTTTAGG	900
	CTATCCTGAA	GCCAAATTG	TCAAGATTGA	ATCTCGGGGA	AAACAGGAAC	AAGAAAGTGT	960
	TATGGTCAAA	GATTTCCCTC	TGGTAACAGA	AACCTCCTTA	GAAGTCTATG	CCAGAGAGGT	1020
10	AGCTGCTTAA	CTAGTGAATA	TTCAAGCTTT	CCAGCAACCG	ATTTTGGTTC	TCITTACCGC	1080
	TAAAGACATG	CTTCTAGCAG	TATCGGATTT	ACTTACAGTT	AGCCACTTGG	CCCAGTATAA	1140
	AAATGGGGAT	GTTTCATCAG	TAAAGAAACG	CTTTGAAAAA	GGTGAACAAC	AAATCTTGCT	1200
15	TGGTCAGCA	AGTTCTGGG	AGGGAGTTGA	TTTTTCAAGC	CATCCTTTTG	TGATTCAAGT	1260
	TGTACCGAGG	CTTCTTTTCC	AAAACTCTCA	AGAACCTTGG	ACGAAAAAGA	TTAATCAAGA	1320
20	ACTGAATCAA	GAAAGGAAAA	ATGCCCTTTA	TGATTATCAA	TTGCCAATGG	CCATTATTCTG	1380
	TTTAAACAG	GCTTTGGGAA	GAAGTATGAG	ACGTGAATAC	CAACGTTCTC	TAACCTTAT	1440
	TTTGGATAG	AGAATCATCG	GAAACGATA	CGGCAACAA	ATAGTAGCAT	CTCTAGCAGA	1500
25	AGAAAGGACT	TTTAAACCA	TCTCTCGATC	CGAAGTTGAC	GAGGCTATTG	ATAGATTTT	1560
	TAATGAACCT	TGATAAATAG	TATTGTATGA	AAGTATAAGG	TTAGTACATA	TGAAACGTTT	1620
30	TCTCGACTCT	AGATCGATT	ATAGTTTCT	CTTGCCAGTA	TTTTTTCTAC	TGTCATCGG	1680
	TGTGGTGGCT	ATCTATATAG	CCGTTAGTCA	TGATTATCCC	AATAATATTC	TGCCCATTTT	1740
35	AGGGCAGCAG	GTCGCTTGA	TTGCCCTGGG	GCTTGTGATT	GTTTTTGTGG	TCATGCTCTT	1800
	TAATACAGAA	TTTCTTTGGA	AGGTGACCCC	CTTCTATAT	ATTTTAGGCT	TGGGACTTAT	1860
	GATCTTGCG	ATTGTATTTT	ATAATCCAAG	CTTAGTTGCA	TCAACGGGTG	CCAAAACTG	1920
40	GGTATCAATA	AATGGAATTA	CCCTATTTC	ACCGTCAGAA	TTTATGAAGA	TATCCTATAT	1980
	CTCATGTTG	GCTGTGTCA	TGTCCAATT	TACAAGAAAA	CATAAGGAAT	GGAGACGCAC	2040
	GGTCCGCTG	GACTTTTTGT	TAATTTTCTG	GATGATTCTC	TTTACCATT	CAGTCTAGT	2100
45	TCTTTTAGCA	CTTCAAGTG	ACTTGGGGAC	GGCTTTGGTT	TTTGTAGCCA	TTTTCTCAGG	2160
	AATCGTTTTA	TTATCAGGGG	TTTCTTGGAA	AATTATTATC	CCAGTATTGG	TGACTGCTGT	2220
50	AACAGGAGTT	GCTGGTTTCT	TAGCTATCTT	TATTAGCAAG	GACGGACGAG	CTTTTCTTCA	2280
	CCAGATTGGA	ATGCCGACTT	ACCAATCAA	TCGGATTTTG	GCTTGGCTCA	ATCCCTTTGA	2340
55	GTTTGCCCAA	ACACGACTT	ACCAGCAGGC	TCAAGGGCAG	ATTGCCATTG	GGAGTGGTGG	2400
	CTTATTGGT	CAGGGATTTA	ATGCTTCGAA	TCTGCTTATC	CCAGTTCGAG	AGTCAGATAT	2460
	GATTTTACG	GTTATTGCAG	AAGATTTTGG	CTTTATTGGC	TCTGCTCTGG	TTATTGCCCT	2520
60	CTATCTCATG	TTGATTTACC	GTATGTTGAA	GATTACTCTT	AAATCAAATA	ACCAGTTCTA	2580

WO 98/26072

PCT/US97/22578

-90-

	CACCTTATATT TCCACAGSTT TGATTATGAT GTTGCTCTTC CACATCTTG AGAATATCGG	2640
	TGCTGTGACT GGACTACTTC CTTTGACGGG GATTCCCTTG CCTTTCATTT CGCAAGGGGG	2700
5	ATCAGCGGATT ATCAGTAATC TGATTGGTGT TGGTTTGCTT TTATCGATGA GTTACCAGAC	2760
	TAATCTAGCT GAAGAAAAGA GCGAAAAAGT CCCATTCAA CGSAAAAAGG TTGTATTAAA	2820
	ACAAATTAAA TAAGGAGAAA ATCATGGTAA AAGTAGCAGT TATGTTAGCT CAGGGCTTTG	2880
10	AAGAAATTGA AGCCTTGACA GTTGTAGATG TCTTGCGTCG AGCCAAATATC ACATGTGATA	2940
	TGGTTGGTTT TGAAGAGCAA GTAACGGGTT CGCATGCAAT CCAAGTAAGA GCAGATCATG	3000
15	TCTTTGATGG AGATTATCA GACTATGATA TGATTGTTCT TCCTGGAGGT ATGCTCGGTT	3060
	CTGCACATTT ACGTGATAAT CAGACCTTGA TTCAAGATT GCAAAGCTTC GAGCAAGAG	3120
	GGAGAAACT AGCAGCCATT TGTGCGGCAC CAATTGCCCT CAATCAAGCA GAGATATTGA	3180
20	AAAATAAGCG ATACACTTGT TATGACGGCG TTCAAGAGCA AATCCTTGAT GGTCACTACG	3240
	TCAAGGAAAC AGTAGTGGTA GATGGTCAGT TGACAACCAAG TCGGGGTCTCT TCAACAGCCC	3300
25	TTGCCCTTTCG CTACGAGTTG GTGGAGCAAC TAGGAGGGGA CGCAGAGAGT TTACGAAACG	3360
	GAATGCTCTCA TCGAGATGTC TTTGGGTAAA AATCAGTAAA ACGGGAGTTA TTCTCTCGTT	3420
	TTTTATGTGG AAAAATCAGG GAAATCATCG CTTTTTTCAT AAAAAAATGC TATAATGAAG	3480
30	GGTATGAAAT ATCACGATTA CATCTGGGAT TTAGGTGGAA CTTTACTGGA TAATTATGAA	3540
	ACTTCAACAG CTGCATTGTG TGAAACATTG GCACGTGATG GTATCACACA AGACCATGAC	3600
35	AGTGTCTATC AAGCTTTAAA GGTTCCTACT CCTTTTSCGA TTGAGACATT CGCTCCCAT	3660
	TTAGAGAATT TTTTAGAAAA GTACAAGGAA AATGAAGCCA GAGAGCTTGA ACACCCGATT	3720
	TTATTTGAAG GAGTTTCTGA CCTATTGGAA GACATTTTAA ATCAAGGTGG CCGTCATTTT	3780
40	TTGGTCTCTC ATCGAAATGA TCAGGTTTTG GAAATTTTAG AAAAAACCTC TATAGCAGCT	3840
	TATTTTACAG AAGTGGTGAC TTCTAGCTCA GGCTTTAAGA GAAAGCCAAA TCCCGAATCC	3900
45	ATGCTTTTAT TAAGAGAAAA GTATCAGATT AGCTCTGGTC TTGTCTTTGG TGATCGGCCG	3960
	ATTGATATCG AAGCAGGTCA AGCTGCAGGA CTTGATACCC ACTTGTTTAC CAGTATCGTG	4020
	AATTTAAGAC AAGTATTAGA CATATAAGAA AAAGGAATAA GATGACAGAA GAAATCAAAA	4080
50	ATCTGCAGGC ACAGGATTAT GATGCCAGTC AAATTCAGT TTTAGAGGGC TTAGAGGCTG	4140
	TTCTGATGGC TCCAGGGATG TACATTGGAT CAACCTCAA AGAAGGTCTT CACCATCTAG	4200
55	TCTGGGAAAT TGTTGATAAC TCAATTGACG AGGCCTTGGC AGGATTTGCC AGCCATATTC	4260
	AAGTTTTTAT TGAGCCAGAT GATTGATTA CTGTTGTGGA TGATGGGCCG GGTATCCCAG	4320
	TGATATTACA GGAATAAACA GGTGCTCTCG CTGTTGAGAC CGTCTTTACA GTCTTTCACG	4380
60	CTGGAGGAAA GTTCGGCGGT GGTGGATACA AGGTTTCAGG TGCTCTTCAAC GGGGTGGGGT	4440



WO 98/26072

PCT/US97/22578

-91-

```

5   CGTCAGTTGT TAATGCCCTT TCCACTCAAT TAGACGTTCA TGTCCATAAA AACGGTAAGA 4500
    TTCAATTACCA AGAATACCCT CGTGGTCATG TTGTCGCAGA TCTTGAAATA GTTGGAGATA 4560
    CGGATAAAAC AGGAACAACCT GTTCACCTCA CACCGGACCC AAAAATCTTC ACTGAAACAA 4620
    CAATCTTTGA TTTTGATAAA TTAATAAAAC GGATTCAGA GTTGGCCTTT CTAAATCGCG 4680
10  GTCTTCAAAAT TTCTATCACT GATAAGCGCC AAGGTTTGGG ACAAAACCAAG CATTATCATT 4740
    ATGAAGGTGG GATTGCTAGT TACGTTGAAT ATATCAACGA GAACAAGGAT GTAATCTTTG 4800
    ATACACCAAT CTATACAGAC GGTGAGATGG ATGATATCAC AGTTGAGGTA GCCATGCAGT 4860
15  ACACAACGGG TTACCATGAA AAATGTCATG AGTTTCGCCA ATAATATTCA TACACATGAA 4920
    GGTGGAACGC ATGAACAAGG TTTCCGTACA GCCTTGACAC GTGTTATCAA CGATTATGCT 4980
    CGTAAGAATA AGTTACTGAA AGACAATGAA GACAATCTAA CAGGGGAAGA TGTTCCGGAA 5040
    GGCTTAACCT CAGTTATCTC AGTTAAACAC CCAATCCAC AGTTTGAAGG ACAACGAAG 5100
    ACCAAATTGG GAAATAGCGA AGTGGTCAAG ATTACCAATC GCCTCTTCAG TGAAGCCTTC 5160
25  TCCGATTTCC TCATGAAAAA TCCACAGATT GCCAAACGTA TCGTAGAAAA AGGAATTTTG 5220
    GCTGCCAAGG CTCGTGTGSC TGCCAAGCGT CCGCGTGAAG TCACACGTAA AAAATCTGGT 5280
    TTGGAAATTT CCAACCTTCC AGGGAAACTA GCAGACTGTT CTTCTAATAA CCCTGCTGAA 5340
    ACAGAACTCT TCATCGTCGA AGGAGACTCA GCTGGTGGAT CAGCCAAATC TGGTCGTAAC 5400
    CGTGAGTTTC AGGCTATCCT TCCAATTCGC GGTAAAGATT TGAACGTTGA AAAAGCAAGT 5460
35  ATGATAAAGA TTCTAGCTAA CGAAGAAATT CGTAGTCTTT TCACAGCCAT GGGAACAGGA 5520
    TTTGGCCGAG AATTGTATGT TTGAAAGCC CGTTACCAA AACTCGTTTT GATGACCGAT 5580
    GCCGATGTCG ATGGAGCCCA CATTGCTACC CTCTTTTAA CCTTGATTTA TCGTTATATG 5640
40  AAACCAATCC TAGAAGCTGG CTATGTTTAT ATTGCCCAAC CACCAATCTA TGGTGTCGAA 5700
    GTTGAAGCG AGATTAAAGA ATATATCCAG CCGGTGCGAG ATCAAGAAAT CAAACTCCAA 5760
45  GAAGCTTTAG CCCGTTATAG TGAAGTCTGT ACCAAACCGA CTATTCAGCG TTATAAGGGG 5820
    CTAGGTGAAA TGGACGATCA TCAGCTGTGG GAAACAACCA TGGATCCCCG ACATCGCTTG 5880
50  ATGGCTAGAG TTTCTGTAGA TGATGCTGCA GAAGCAGATA AAATCTTTGA TATGTTGATG 5940
    GGGGATCGAG TAGAGCCTCG TCGTGAGTTT ATCGACTCTA GAGGATCCCC GG 5992

```

55 (2) INFORMATION FOR SEQ ID NO:40:

```

    (1) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 907 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
60  (D) TOPOLOGY: linear

```

WO 98/26072

PCT/US97/22578

-92-

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

5 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

TACAAAAGTA	GGTGGAGAGG	CTGATTATTT	GGTCTTTCCA	CGAAATCGTT	TTGAGTTGGC	60
TCGCGTTGTG	AAATTGCGCA	ACCAAGAAAA	TATCCCTTGG	ATGGTCTTGG	GCAATGCAAG	120
CAATATCATC	GTTGCTGATG	GTGGGATTCC	TGGATTGTGC	ATCTTGTGTG	ACAAGCTCAA	180
TAAAGTTTTCT	GTTGATGGCT	ATACCATTGA	AGCAGAAAGCT	GGGCTAACT	TGATTGAAAC	240
20 AACTCGCATT	GCCCTCCGTC	ATAGTTTAAC	TGGCTTTGAG	TTTGCTTGTG	GTATTCCAGG	300
AAGCGTTGGC	GGTGCTGCTT	TTATGAATGC	GGTGCCTTAT	GGTGGCGAGA	TTGCTCACAT	360
CTTGCACTCT	TGTAAGGTCT	TGACCAAGGA	TGGAGAAATC	GAACCCCTGT	CTGCTAAAGA	420
25 CTTGGCTTTT	GGTTACCGCC	ATTGAGCTAT	TCAGGAGTCT	GGTGCAGTTG	TCTTGTGAGT	480
TAAATTTGCC	CTAGCTCCAG	GAACCCATCA	GTTTATCAAG	CAGGAAATGG	ACCGCTTGAC	540
30 GCACCTACGT	GAACCTAAGC	AACCTTTGGA	ATACCCATCT	TGTGGCTCGG	TCTTTAAGCG	600
TCCAGTCGGG	CATTTTGTCAG	GTCAGTTAAT	TTCAGAAAGT	GGCTTGAAAG	GCTATCGTAT	660
35 CGGTGGCGTA	GAGTGTCTAG	AAAAGCATGC	AGGATTATAT	ATCAATGTCT	CAGATGGAAC	720
GGCCAAAGAC	TACGAGGACT	TGATCCAATC	GGTTATCGAA	AAAGTCAAGG	AACACTCAGG	780
TATTACGCTT	GAAGAGAAAG	TCCGGATCTT	GGGTGAAAGC	CTATCGGTAG	CGAAGATGTA	840
40 TGCAAGGTGT	TTTACTCCCT	GCAAGAGGTA	GTGGGGACCT	GACAGAGCCC	CGATCGGTTA	900
AGCTATG						907

(2) INFORMATION FOR SEQ ID NO:41:

45 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2764 base pairs

(B) TYPE: nucleic acid

50 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

55 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

	AGAACCCTTG	GATGCAGCCA	TTCAAGAAGT	TTCTCCAGAA	TTGTTTGACC	AATATGAAAT	60
5	CTTTAAATCA	CGTGAATATG	TGCTAGAAATG	GTCCACAAAG	AATGTTTATA	AAGCAACAGG	120
	TTTGCCAAAA	CTAATCAGCC	ATCTTGGAAT	CGACCAAGT	CAAGTGATGG	CTTGTTGGTGA	180
	CGAGGCCAAT	GACCTCTCTA	TGATTGAATG	GGCAGGTCTT	GGTGTGCTA	TGCAAAAACGC	240
10	TGTTTCTGAA	GTAAGGCGAG	CCGCAATATG	AGTGACGCCG	ATGACCAACG	ATGAGGAAGC	300
	TGTCGCCTGG	GCTATCGAAG	AATATGTGCT	AAAGGAGAAC	TAAGATATGG	GATTGTTTGA	360
	CCGTCTATT	GGAAAAAAG	AAGAACCTAA	AATCGAAGAA	GTTGTAAAA	AAGCTCTGGA	420
15	AAATCTTGAT	TTGTCTGAAG	ATGTTGATCC	TACCTTCACA	GAAGTTGAGG	AAGTTTCTCA	480
	GGAAGAAGCA	GAGGTTGAAA	TTGTTGAACA	AGCTGTGTTT	CAAGAAGAGG	AAATCCAAGA	540
20	CACAGTTGAA	GAAAGTCTGG	ATTTAGAGCC	AGTTGTAGAA	GTTTCTCAAA	AAGAAGTCGA	600
	AGAATTTCCA	CACCTCAGAAG	AAGGGAATAC	TGAGTTTCTA	GAGACTATAG	AAGAAAATAA	660
	TTCTGAAGTT	CTTGAACAGC	AAAGGCCTCA	AGCAGAAGAA	ACCGTTACAG	AAAAATATGA	720
25	CCGCAGTCTT	AAGAAAACCT	GTACAGGTTT	CGGTGCCCGC	TTGAATGCCCT	TCTTTGCTAA	780
	CTTCCGCTCT	GTTGACGAAG	AATTTTTCGA	GGAAGTGGAA	GAACTGCTGA	TTATGAGTGA	840
30	TGTTGGTGTC	CAAGTCGCTT	CTAACTTAAC	GGAGGAACTA	CGTTACGAAG	CCAAGCTTGA	900
	AAATGCCAAG	AAACCTGATG	CACCTTCGTCG	TGTCATCAT	GAGAAATTGG	TTGAGCTTTA	960
35	TGAAAAGGAT	GGTAGCTACG	ATGAAAGCAT	CCACTTCCAA	GATAACTTGA	CAGTTATGCT	1020
	CTTTTGGTGG	GTGAATGGTG	TTGGGAAAC	AACTTCTATC	GGAAAACTAG	CCCACCGCTA	1080
	CAACCAAGCT	GGTAAGAAGG	TCATGCTGGT	TGCAGCAGAT	ACCTTCCGTC	CGGGTGCAGT	1140
40	AGCTCAGCTA	GCTGAATGGG	GCCGACGAGT	AGATGTTCCA	GTAGTAACCT	GACCTGAAAA	1200
	AGCTGATCCA	GCCAGCGTGG	TCTTTGATGG	TATGGAACGT	GCCGTGGCTG	AAGGTATCGA	1260
	TATTCTCATG	ATTGATACTG	CTGGTCTGCT	GCAAAATAAG	GATAACCTTA	TGGCTGAGTT	1320
45	GAAAAAGATT	GGTCGTATTA	TCAAACGCTG	TGTGCCAGAA	GCACCACATG	AAACCTTCTT	1380
	GGCACCTTAT	GCATCAACAG	GTCAAAATGC	CCTAGTACAG	GCCAAAGAA	TTTCGAAAA	1440
50	CACACCTTTA	ACGGGAATTG	TTTGTACTAA	GATTGATGGA	ACTGCTCGAG	GAGGTCTGGT	1500
	TCTAGCCATT	CGTGAAGAAC	TCAATATTCC	TGTAAAATTG	ATTGGTTTGG	GTGAAAAAAT	1560
55	CGATGATATT	GGAGAGTTTA	ACTCAGAAAA	CTTTATGAAA	GGTCTCTTGG	AAGGTTTAA	1620
	CTAATCAGAA	GCAAAAATCC	TGCAAGGCAT	AAACTTGACG	GAAATTTTTT	TATTCTAAGC	1680
	GACCATCTTG	ACGATAGGTT	ATATCTGGTT	GCCAAGTCCA	TTTGGCACCG	AATTTTTCAA	1740
60	GTAGGTCAAA	GCTGGCTTGA	GGTCCCATGC	TTCCAGCTTT	ATAGTCATGA	AGTGGGGCAC	1800



WO 98/26072

PCT/US97/22578

-95-

	AGACTAGTAG AATTAGTCAT TAGATATAGA ATTTTAGTGA GTCCAAAAGA TGTTCAAAAGA	300
	TTTATCCAA TCTATTATTG AAAAGTTTA TATTATAAT TTCGAAAAAT GCTCTCAAT	360
5	ACCCTGTTTG ACGAGTGAAG AATTGAAAGT CTTGGAAAAA GGTATGTCCT GACTGGTAAA	420
	GAATGGATT GTCAATCAGA TGATGAGCTG GAAGAATTTA AAAATCTATT TTTAAATTTT	480
10	ATCAATCCTG AAGAATGGGA TACTATCTCC TTTGATTGAG ATTTTATGCC GTTTCACCAA	540
	TCGTAACCAA TTTCTCAAAA AAGTTAAATC TTATATTAG TACTCTGTAA AACTCTTATC	600
	TAATCACGTT GCTTATACTC AATGAAAAAT ATAGAAAAAA GCATAGTATC AGGTGTTGAA	660
15	ACACCTGATA CTATGCTTTT TATTGTGGGA AGATTACTT TTTTCTTCTT GAAATTGAGT	720
	TGTTACCCAG GCTCTTTCAG TTATTAAGC CTTGATGACT TTAATGTGTT TAGATAGCTT	780
20	AAAAGGATT GAATCACTTA GTTTAGAATC TGAACAATA GTATCAAGAT TTGATATATT	840
	ATAAAAAGTA TAAAAATCAA ACTTATTGAA CTACTATGTA TCTGCGAGTA AATATTTTTT	900
	ATTAGAATTA TTTAAAGCGA TCGGTTGAGC CTCTCCCTCT TCCTCGCTAA AAGTAGCTAG	960
25	AGCTCCGTTT TGAATACCAT TACAGCTAAC GAAAGCTTTA GAAATTTGGA GATTAGAGAG	1020
	ATTTTGTAGG GTCAATGTGC CAACAAAAGC ACCTGTAATA TCGCGATAAT TTCCACCTAT	1080
30	CAAAATCAAA CTCGTTAATT TTCGTTTCGT TAAATCAGA AAACAGGTA GACTGTTGCT	1140
	TACGACGCGG ATATTGTCAA TAGGCAACTC ACGCGCAAAA AACTCTAATG TTGTTCTCTG	1200
	TCCAATGAAA ATAGTTTCTC TTTCTTCTAC TAGACTGCCT GCAAAATGGG CTATTTCCTG	1260
35	TTTTTCTGCC GTTTGGAGGG CTGTTTTTC AATATTGAT CGCTCATAG TCAAAAGGGA	1320
	GTTGGTTCGA AGTTTTTCAG CTCCACCATG CACACGAATC AGCAATCTT TATCAGCTAA	1380
40	TTCTGTAAA TAGCGCCTTG CAGTCATATC TGAAACGGCT ATTTCTGTCA TAATCTGTTT	1440
	AACGTGTTAT GTTCCTTTAC TATTACTAT CTCTAAATTT TTGGCTAATT TTTCTGTGTT	1500
	GAGCATATTA TCACCTCGTT TCCTACTACT ATCTTACCAT AAACAAACTC ATCATTCAAA	1560
45	TACAAAAACA ACAAATGAA ACAAAAACAA AAATATCGAA GTTTGTTTTT AAACTTTTCG	1620
	ATATTTTGTG TGGGTTATAA CTTTGATGTT TCTAGTTTAC TTTTGTATGA TTGAGAGTGA	1680
50	TGGAGAATTA GTCTAAACCG TAGTTATAGT CATCGCTTGG CATGGCTTCA ACTTCGCCAA	1740
	GAAGGTAAAC ATTTCCGACT TGAGAGAAGA AGTCATGTTT GGAAGTTCCT GTTGAAATAC	1800
	CGTTCATAAC GATTGGGTG ACATCTTCAG CTGAATCTGG GAAAGTGGGA TCTTGTCCEA	1860
55	TGTTCAATGAG AGCTTTATTG GCATTGTAGC GAAGGAAGGT TTTAACCTCT TCAGTCCAAC	1920
	CAACACCGTC ATAAAGACTC TCTGTGTAGC CTTCTTCATT TTCATAAGA GTATAGAGTA	1980
60	GGTCGTACAT CCATTCTTTG AGTTTTTCTT GCTCTTCTTC AGGCAATTCA TTGAAACCAA	2040
	GTTGGAATTT GTAACCAATG TAGGTTCCGT GAACGAGACTG GTCACGAATA ATCAATTTAA	2100

WO 98/26072

PCT/US97/22578

- 96 -

	TGATTCTGCG AACGTTGGCT AGTTTGTGTT TACCGAGATA GTAGAGGGGA GTGAAGAAAC	2160
5	CAGAGTAGAA GAGGAAGGTT TCGAGGAAGA CGTCGGCAAC TTTCTTTTCA AGTGGGCTGC	2220
	CGTTTAGGTA GATTTCGTTG ACAATCTCAG CCTTCTTTTG TAGGTAAGGA TTGGTATTGG	2280
	TCCATTCGAA AATTCTTCTA ATCTCAGCCT TAGTATTCAA GGTAGAAAAG ATTGATGAGT	2340
10	AAGATTTAGC GTGGACAGAT TCCATAAATT GGATGTTAIT GAAAACAGCT TCCTCATGTG	2400
	GTGTACGGAT GTCTGCGCGA AGGGCTTGAA CCCCAGTTTC AGATTGCATA GTGTCAAGAA	2460
15	GGGTTAAACC ACCAAAAAAT TTTCCGACCA AGTCTTTCTC TTTGTTAGAT AGCTTCTTCC	2520
	AGTCATCCAA GTCGTTTGAT AAGGGAATAC GTGTATCGAG CCAAAATTGC TCCGTCAAGT	2580
	TTTCCCAAGT TGATTGTGCG ATGACATCTT CGATGGCAIT CCAGTTAATG GCTTTGTAGT	2640
20	AAGTTTCCAT TTAANAATCT TTTCTGTGTT TAGTATTGCG AACTCACAAT TATTCTACT	2700
	TTACCATAAT TCTATAGGAG TATCGCACAA AAAGTCGGAA GCCCGACTTT TAAAAATGTA	2760
25	CATAAATAT GTTATGACAT AGTAGATTG ATTTTATCAG TGCTGCTTAG GAAAAATPAA	2820
	TGTTTCTATG CTAGAAACTA AATCACACAG CTTTCACATT GGTGGCGCC GACTTCTCCA	2880
	CGTCATCTG TAAAGGTACG GACGTAGTAG ATAGACTTGA TTCCCTTGT AAAGGCATAG	2940
30	TTACGAAGGA TGGACAAATC ACGTGTCGTT TGTTTATTTT CCTCTTCCA TTCGTAAAGG	3000
	CCTTTGGAA TGTCACTACG CATGAAGAGG GTGAGTGAAA GTCCCTGATC CACGTGTTCA	3060
35	GTGCGACGAG CGTAAACATC GATGACTTTA CGCATATCCA TATCGTAGGC AGAAGTGTAG	3120
	TAAGGAATGG TTTCTGTAGA CAAGCCAGCA GCAGGGTAGT AGATTTTACC AATTTTCTTC	3180
	TCTTGGCGCT	3189

40 (2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3580 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

50 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

60	TTATTGAAGA AGGTGTTAAA GTTGTACCAA CAGGAGCAGG AAATCCAAGC AAGTATATGG	60
	AACGTTTCCA TGAAGCTGGG ATAAATCGTTA TTCCTGTGCT TCCTAGTGTC GCTTAGTCTA	120

PCT/US97/22578

	AACGCATGGA	AAAATCGGTT	GCAGACGCTG	TTATTGCAGA	AGGAATGGAA	GCTGGGGGGC	180
5	ATATCGGTAA	ATTACAAACC	ATGACCTTGG	TGCGACAGGT	AGCCACAGCT	GTATCTATTCT	240
	CTGTTATTGC	TGCAGGAGGA	ATTGCGGATG	GTGAAGGTGC	TGCGGCTGCG	TTTATGCTAG	300
	GTGCAGAGGC	TGTACAGGTT	GGGACACGGT	TTGTAGTTGC	AAAAGAGTGC	AATGCCCATC	360
10	CAAACTACAA	GGAGAAAATT	TTAAAGCAA	GGGATATTGA	CACTACGATT	TCAGCTCAGC	420
	ACTTTGGTCA	TGCTGTTCTG	GCTATTAAAA	ATCASTTGAC	TAGAGATTTT	GAACCTGGCTG	480
15	AAAAAGATGC	CTTTAAGCAG	GAGATTCCTG	ATTTAGAGCT	CTTTGAACAA	ATGGGAGCAG	540
	GTGCCCCIAG	CAAAAGCATT	GTTTCAACCG	ATTGTGAGGG	TGGCTCTGTC	ATGCGCAGTTC	600
	AAATCGCAGG	GCTTGTTTCT	AAAGAAGAAA	CAGCTGAAGA	AATCCTAAAA	GATTTGTATT	660
20	ACGGAGCCGC	TAAAGAAAAT	CAAGAAGAAG	CCTCTCGCTG	GACAGGAGTT	GTAAGAAATG	720
	ACTAAACAGC	CCTTTTATT	TGCTGGTCAA	GGTCCCAGT	ATCTAGGGAT	GGGACGGGAT	780
25	TTCTATGATC	AGTATCCGAT	TGTTAAAGAA	ACGATTGATC	GAGCGAGTCA	GGTGCTAGGT	840
	TATGATTTCG	GTTATCTCAT	CGATACGGAA	GAAGACAAC	TCATACAGAC	CCGCTATACG	900
	CAACCAGCCA	TTCTAGCGAC	TTGCGTTGCT	ATCTACCGTT	TATTGCAAGA	AAAGGGCTAT	960
30	CAGCCTGATA	TGGTTGCTGG	TTTGCTCTTT	GGAGAATACT	CTGCCTTTGT	GGCAACGGGC	1020
	GCCTTGAGATT	TTGAAGATGC	GGTTGCCCTT	GTAGCTAAGC	GTGGAGCCCTA	TATGGAAGAA	1080
35	CGCGCTCCTG	CTGACTCTGG	CAAGATGGTA	GCAGTCTCTCA	ATACGCCGAT	AGAGGTCATT	1140
	GAGAGAGCCT	GTCAAAAAGT	TGCTGAACCT	GGAGTGGTTA	CTCCAGCCAA	CTATTAACACA	1200
	CCTGCACAAA	TCGTCAITGG	TTCGAAGAAG	TTGTGCACTG	ATCGACGGGT	TGAACITTTT	1260
40	CAAGAAGCAG	GTGCCAAACG	CTTGATTCTC	CTTAAGTGTG	CAGGTCCCTT	TCACACCTCT	1320
	CTCCTTGAAC	CTGCTAGCCA	GAAACTAGCT	GAAACTCTGG	CTCAGGTAAG	TTTTTCAGAT	1380
45	TTTACTTTGC	CCCTAGTCGG	CAATACAGAA	GCTGCTGTGA	TGCAAAAAGA	GGACATTGCT	1440
	CAGCTCTTGA	CGCGTCAGGT	CAAGGAACCC	GTTGCTTTCT	ATGAAAGTAT	TGGGGTCATG	1500
	CAAGAAGCAG	GCATAAGCAA	CTTTATCGAG	ATTGAGCCGG	GGAAAGTCTT	GTCAGGTTTT	1560
50	GTAAAAAAAT	TTGATCAAA	TGCTCACTTA	GCTCATGTGG	AAGATCAAGC	GAGTTTAGTA	1620
	GCACTTTAG	AAAATAAGAC	TAAATAAGT	AGAAGTTTGT	AAAGGAAAAA	AATGAAACTA	1680
55	GAACATAAAA	ATATCTTTAT	TACAGGTTGC	AGTCGTGGAA	TTGGTCTTGC	CATCGCCAC	1740
	AAGTTTGTCT	AGACGAGGAC	CACATATTGC	TTAAACAGTC	GGTGGGCAAG	CTCAGAGTAA	1800
	TGTGCTGCTG	AGTTTTCAAA	TACTGGTATC	AAGGTGGTCT	CCATTTCAGG	AGATGTGATCA	1860
60	GATTTTGCAG	ACGCTAAGCG	TATGATTGAT	CAGCAATTG	CAGCAATGGG	TTTCAGTAGAT	1920

WO 98/26072

PCT/US97/22578

-98-

	GTTTGGGTCA	ACAATGCAGG	GATTACCCAA	GATACTCTTA	TGCTCAAGAT	GACAGAAGCA	1980
	GATTTTGAAA	AAGTGCTCAA	GGTCAATCTG	ACTGGTGCCT	TTAATATGAC	ACAATCAGTC	2040
5	TTGAAACCGA	TGATGAAAGC	CAGAGAAGGT	GCTATCAITA	ATATGTCTAG	TGTTGTGGT	2100
	TTGATGGGGA	ATATTGGTCA	AGCTAACTAT	GCTGCTCTTA	AGGCTGGCTT	GATTGGCTTT	2160
	ACCAAGTCTG	TGGCAGCGGA	GGTCGCTAGT	CGGAATATAC	GAGTCAATGT	GATTGTCCCA	2220
10	GGAAATGATTG	AGTCTGATAT	GACAGCTATC	TTATCAGATA	AGATTAAAGGA	AGCTACACTA	2280
	GCTCAGATTG	CGATGAAAGA	ATTTGGGCAG	GCAGAGCAGG	TTGCAGATTT	GACAGTATTT	2340
15	TTAGCAGGCC	AAGATTATCT	AACGTGCTCA	GTGATTGCCA	TTGATGGTGG	CTTAAGTATG	2400
	TAGCGAAAGC	TAGAGGTGAA	AAGAATGAAA	CTAAATCGAG	TAGTGGTAAC	AGGTTATGGA	2460
	GTAACATCTC	CAATCGGAAA	TACACCAGAA	GAATTTTGGA	ATAGTTTAGC	AACTGGGAAA	2520
20	ATCGGCATTG	GTGGCATTAC	AAAATTTGAT	CATAGTGACT	TTGATGTGCA	TAATGCGGCA	2580
	GAATCCAAG	ATTTTCCGTT	CGATAAATAC	TTTGTAAGAA	AAGATACCAA	CCGTTTGTAT	2640
25	AACIATTTCT	TATATGCTT	GTATGCAGCC	CAAGAGGCTG	TAAACAGGCC	AATCTTGATG	2700
	TAGAGGCTCT	TAATAGGGAT	CGTTTGGTGG	TTATCGTTGC	ATCTGGTATT	GGTGGGAATCA	2760
	AGGAAATTGA	AGATCAGGTA	CTTCGCCTTC	ATGAAAAAGG	ACCCAAACGT	GTCAAACCAA	2820
30	TGACTCTTCC	AAAAGCTTTA	CCAAATATGG	CTTCTGGGAA	TGTAGCCATG	CGTTTGGGTG	2880
	CAAACGGTGT	TTGTAAATCT	ATCAATACTG	CTTGCTCTTC	ATCAAATGAT	GCGATTGGGG	2940
35	ATGCGCTCCG	CTCCATTAA	TTTGGTTTCC	AAGATGTGAT	GTGGTGGGA	GGAAACAGAG	3000
	CTTCTATCAC	ACCTTTTGCC	ATCGCTGGTT	TCCAAGCCTT	AACAGCTCTC	TCTACTACAG	3060
	AGGATCCAAC	TCGTGCTTCG	ATCCCATTTG	ATAAGGATCG	CAATGGGTTT	GTTATGGGTG	3120
40	AAGGTTCAGG	GATGTTGGTT	CTAGAAAGTC	TTGAACACGG	TGAAAAACGT	GGAGCTACTA	3180
	TCCTGGCTGA	AGTGGTTGGT	TACGGAATA	CTTGTGATGC	CTACCAATGC	ACTTCTCCAC	3240
45	ATCCAGAAGG	TCAGGGAGCT	ATCAAGGCCA	TCAAACCTAGC	CTTGGAAAGAA	GCTGAGATTT	3300
	CTCCAGAGCA	AGTAGCTATG	TTAATGCTCA	CGGAACGTCA	ACTCCTGCCA	ATGAAAAAGG	3360
	AGAAAGTGGT	GCTATCGTAG	CTGTTCTTGG	TAAGGAAGTA	CCTGTATCAT	CAACCAAGTC	3420
50	TTTTACAGGA	CATTTGCTGG	GGGCTGCGGG	TGCAGTAGAG	CTATCGCACC	ATCGAGCTAT	3480
	GGCTCATACT	TTGTACCATG	CCAGCTGGGG	AAGTGAGGTA	TCAGATATAT	CGAGCTAATG	3540
55	TCGTTATGGC	AGGTTTGAGA	AGAATTCATA	CGTATTCAAA			3580

(2) INFORMATION FOR SEQ ID NO:44:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1780 base pairs  
 (B) TYPE: nucleic acid



-99-

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

15	ATGTCCCGCAA GAATGTGATT AATCAGCAAT CATCCTTGAT CGGAGATGAA TCGATCTTGG	60
	CTTTTGGAGT GAACCAGCCT TTTAGCGGAT TTGGTGTAA AGGAGAAAAG CAGCAACAGC	120
20	ATCAGCCTAT GACCTCTATG TACTGAAGCG ACCACTTCCC CAAGTAGGAC CTCGATGTCA	180
	TTTGTAGATAG TCAAATCAG GCTGCTGCA TTGTCGAAAT TACAAAGGTT TCTGTGAAC	240
	TCCTCAATCA AGTTTCTGCG CAACATGCCT TTAAGGAAGG TGAGGGAGAC AAATCACTTG	300
25	CCTATTGGCG CCAGGTTTAT GAGGACTTTT TCACAGACTG TTTGGGTGAA GTAGGGCTGA	360
	CTTTTACACC TGAAGCAAGG GTTGTTTTAG AAGAAATTCG CAAGGCTTAC CCACGTAGA	420
	CTATTAGAAG GAAGAAAGTT TTGGAATCG CTGTCCAATC CTTTTTCTC AAGCAAAATA	480
30	TGATATAATA AGTTTGTGTT AAGAAGAGCA GCAGCTCTTA AACTTAGAAT AGGAGAAAAC	540
	TATGCAAGCA GTTGAACATT TTATTAAGCA ATTTGTTTCT GAACATTATG ATTTATTTTT	600
35	AGATTGAGT CGTGAGACCA AGACTTTTTT TGGGAAAGTG ACCATCACTG GTCAGCACCA	660
	GAGTGACCGC ATCTCCCTCC ACCAAAAAGA CTTGGAATC ACCTCTGTAG AAGTTGCAGG	720
40	TCAAGCTCGT CCATTTCAGG TTGACCATGA CAATGAAGCC CTTCATATCG AATTGGCTGA	780
	GGCTGGTCAA GTTGAATTGG TTCTTGCTT TTCTGGTAAA ATTACAGACA ACATGACAGG	840
	GATTTACCCT TCTTATTATA CAGTTGATGG AGTCAAGAAG GAGGTCTTGT CTA CTCTCAGTT	900
45	CGAGAGCCAT TTTGCGCGCG AAGCTTTCCT ATGTGTGGAT GAGCCTGAAG CCAAGCAAC	960
	TTTGTACCTC TCTCTTCGCT TTGACCAAGC AGAAGGTGAA TTGGCCTTGT CAACATGCC	1020
	AGAAATCGAT GTTGAAACC GTAAGGAAAC AGGTATCTGG AAGTTTGAGA CAACACCTCG	1080
50	CATGCTCTCT TACTTGTGG CCTTGTGTGC TGGTGATTTG CAAGGGGTGA CCGCTAAAAC	1140
	TAAAAATGGT ACCCTGGTAG GTGTCTACTC AACCAGCA CATCCACTTT CAATCTTGA	1200
55	TTTCTCACTG GATATCGCTG TTGCTCTAT CGAGTTTAC GAAGTACT ATGAGATTAA	1260
	GTACCCAAAT CCTCAATCTC TCCACATCGC CCTTCCTGAC TTCTCAGCTG GTGCTATGGA	1320
	AAACTGGGGT CTTGTGACCT ACCGTGAAGT TTA CTGTGTT GTGATGAGA ACTCTACATT	1380
60	TGCTAGCCGT CAACAAGTTG CCCTTGTGTG GGCCCATGAA TTGGCTCACC AATGTTTGG	1440

WO 98/26072

PCT/US97/22578

-100-

GAACCTCGTG ACTATGAAAT GGTGGGATGA CCTTTGGCTC AATGAAAGTT TCGCTAATAT 1500  
 5 GATGGAATAC GTCTGTGTGG ATACCATCGA ACCAAGCTGG AATATCTTTG AAGATTTCCTA 1560  
 AACAGGTGGA GTACTCTTTG CTCTTGAACG TGACGCTACT GATGGCGTTC AGTCTGTCCA 1620  
 CGTCGAAGTT AAACATCCAG ATGAGATCAA TACACTCTTT GACGGCGCTA TCGCTATATGC 1680  
 10 AAGGAAGCGT CTCATGCACA TGCTTCGCGT TGCTAGAGAT GCTGATTGTG AAGGTTGCAC 1740  
 GCCTACTTTG GAAACACCATT ACAGCACACC ATTGGAGTGA 1780

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 671 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 20 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GTCTTTTSTA GCGAGGCCAG TGTCTTTTGC CCATCATTG TCAGGCAGAT AAAACTAGAG 60  
 35 CGTCTATCTT GATGGCAACA CATGCGACTG AGTAGACCGC AATTTTATAG TTCCAAGCGA 120  
 GCCACCATCC TAGAACTGC GCTCGGGCTC AGATGAAGCT TATCTGGCAG GTCAATCTGG 180  
 40 CGTAGAGATT TTTCTTCAGC CAAGTCCAGA TAGTAGAGCA GGTAGAAGCT TTTCAAGGTC 240  
 AGACTTTGCT CGCTCTGTTG GGCAATGGTC TCTTCCAAGA GACTTCAAT TTCTTTCTGA 300  
 CGCCGATGTA AGTCAAACCA TTTTCCAAA TAGGTCATAG TGTCTCCTT CTTTTTAGAG 360  
 45 TCATAAAATA GAAGAAAGTC CATTAAACGGG CAGTCTCTGC GTCAACAAGT GATTGCGCAT 420  
 GCAATAATTA TACTACTTTT CAAGAATGCT GGCAAGCTCT GTTTTTTAGT GGTTTTCTTT 480  
 50 TTTACTGTCT ATATTTTGG TAAAAATCAA CTTTACTCTG GATGAAGGTT TTGGCTTCAC 540  
 GTAGGAGTTG AAGAAGGGTG GCGCGGGTTT CAATTCCTCT CTGTCTTGG GCAGACTGCG 600  
 GTTCGGGAAG ACTTCCAGAT AACGTTCAAT TTCATCTAG AATCAGAGCA GGATTGGTCT 660  
 55 GGCTCAGTGA C 671

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1557 base pairs  
 (B) TYPE: nucleic acid

WO 98/26072

PCT/US97/22578

-101-

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

15	TTTCAGCTCA CAATATAGG TCGGATGAGC CACTTCCTTA CGAACACGCG CATCAAAAGC	60
	ATCTAGTCCC TCACGTGAAA AAGCATCCTG CAAACTATAA AGAGGATACT GATGACTGTA	120
	TTTTTCAAAA CCATCTAAAA CCTTGCCACC AACACGATGA GTCGGACTGT CTGCTAGCAC	180
20	TTGCTCTGGA TAAGCAGTTT CTAACTCGAC CAACTCACGG TAAAGGCGGT CATACTCACT	240
	GTCTGAAACC GAGGGATTAT CGCTGGTATA GTACTCAGTC GCATAGCGAT TGAGCAAAGC	300
25	GACTAACTCA TTCATTCTTT TATTCATAAG ACCATTTTAC CATAAAACAA GCCCTCCTCA	360
	CAACAGGAAA GGCGGAAAA AACACTTAGT TTGAAATTAT TTTTGAACT CAAGCAACCT	420
	TATATCAATT TTTCAAAATG AGTTCGAACA TAAATAAAGC ATATACAAGA CAAGATGATA	480
30	ACACCCTTC CAATTATCAG GAAAGAAGAG AGATGTACAC TTGGCAAGAC TGTATAAAT	540
	CCTTTTGCAA TAGGCATAAA TAGAATAGCT AAGGTAAAAA TTGTACTCAG TACTCTCCA	600
35	AGAAATTCGC TCTCAACCTT GGTGTGACT TGAGTAAAAA AGTGAATATT AAAAACTGC	660
	ATAACAATT CACAACTAA ATTTCCAGAA AAGGAAAGAA AAGTTGGAAG TGGTAATCCC	720
	ATCATAAAAA CTCGACACC TGTCAAAGCC AGTAAATCA AAGATTATA AATATTAGCT	780
40	TTAATTTTAC TAGCTAGAAG AGCCCCAATG ATGGAACCAA TAGCCCCCAT AGTTAAAATA	840
	CTTGCATAGG CTCCTTCTGA CCCGTAAAGC TGATTCGAAA AGGGAAGTAG AAATTCAAAA	900
45	GCTGCAAAAA AGAAATTAAC GCTGGAAGCT ACCAGCAAAA GGAAGAAAAA TTCTTGCTGA	960
	TGCCAGATAT AGTGTAACCC ATCCTTGATA TCTACAAAAA TATCTCTCCC AGTAAAGGCC	1020
	TTTTTCTCTT GAACCTTTGC TTCCTCTTTT GGAAGGAAG CCACTAGAAC AAAAGCAATG	1080
50	AAAAAAGTCA GCGAGCTAG CAGTAGCGTC ATATGGAGAC TTGCAAACTG TAAAAAAGG	1140
	AAGGAAAGAA CAGGAGAGCT AACACCTACA ACCTGCAAAA CCAGCTCTAA GCGAGAATTA	1200
55	TAGATCACAA TCTCATTTTT CTCCACCACT TCAGTTATGA TAGCTTTATT GGCTGTGCGA	1260
	GAAAGGCGAA AAGCAATAGC CTGCACAATG TTAGCAACAA TCAAGCGGCC AATCATCCAG	1320
	CTATCATTCCT TTATGAAAGA AATAGCCAGA CAAAGAACTC CACAAACAAG ATCTGCCGTC	1380
60	ATTAATAATCT TACGACGAGA AAAACGGTCT GAAATAACTC CGCCAAAGGG ATTGACGAGA	1440

WO 98/26072

PCT/US97/22578

-102-

ATAGATGTGA CGAGCTCAGA AATCTGATAC ATTCCTAAAA CTGCTGTGCC TATAGTCCCC 1500

5 ATAGAAGCCA ACCAGACACT ATTTCCATAA TCATAGAGCA TATCCCATT TTATTGA 1557

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 658 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

25 CTTATTTGGT TTGGGAATTC GTCATGTCGG AAGCAAGGCT AGTCAGCTTT TACTTCAATA 60

TTTCCATTCA ATTGAAAATC TGTATCAGGC AGATTCAGAG GAAGTGCGTA GTATTGAAAG 120

TCTAGGTGGC GTGATTGCCA AAAGTCTTCA GACTTATTTT GCGGCAGAGG CACTCTGAAAT 180

30 TCTGCTCAGA GAATTGAAAG AAAGTGGGGT CAATCTGGAC TATAAAGGAC AGACGGTAGT 240

AGCGGATGCG GCCTTGTCAG GTTTGACCGT GGTATTGACA GGAAAATTGG AACGACTCAA 300

35 GCGCTCAGAA GCTAAAAGTA AACTCGAAAG TCTGGGTGCC AAAGTGACAG GTAGTGTTC 360

TAAAAGACC GACCTCGTCG TGGTAGGTGC AGACGCTGGA AGTAACTGC AAAAAGCACA 420

40 AGAACTTGGT ATCCAGGTCA GAGATGAGGC ATGGCTAGAA AGTTTGTAAT GGATCGTTTA 480

AAAACAGAGT TTAGAGAATA TGAATATGTC TGTAAATTGA GACGAGATTG ACAAATAATT 540

ATTAGTGAAG TAGGAAACAA AGTAAAAGG AAAAATAAAA AATGTATACT ACCCTATGCG 600

45 CATTCATTAC CATCGTAAGA ATGGAGAATA TGACCTTGCT CCTTTGTAAA AGTCAGGA 658

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2474 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

PCT/US97/22578

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

5	ACAATGCATC	AGACAGTCAA	TCGATTCTTA	AAATGTTTAG	AGTAGAGATG	TACCTATTCT	60
	AGTTCAATAT	ACTATATAAC	TGAAUAATT	GATAAAATTAG	TTTTGGAAAT	GACTTAACCAA	120
10	AGATATCCAA	AGTAGTCTAA	AATTGTCTAT	ACTTTATGAG	TGTTTTAGTT	AGGAAAAAGG	180
	CTGTGTGCT	ATAATTGGCG	CATTAGTCTA	GATTTTATTT	ATAGAAAAAT	TTATAATAGA	240
	CTGTATTTAA	AAAATTTTAA	GGAGAAATGA	CAGAATGTCT	GTATCATTTG	AAAACAAAGA	300
15	AACAACCGT	GGTGCTCTGA	CTTTCACAT	CTCTCAAGAC	CAAAATCAAC	CAGAATTGGA	360
	CGGTGTCTC	AAGTCAGTGA	AGAAATCTCT	TAACTTTCCA	GTTTTCCGTA	AAGGTCAACT	420
	CCGAGGCGCT	ACTTCTCGAC	AAAAATTTTG	TGAAGAAGCT	CTTTATCAAG	ATGCAATGAA	480
20	CGCACITTTG	CCAAACGCTT	ATGAAGCAGC	TGTAAGAAAG	GCTGGTCTTG	AAGTGGTTGC	540
	CCAACCAAAA	ATTGACGTAA	CTTCAATGGA	AAAAGGTCAA	GACTGGGTTA	TCACTGCTGA	600
25	AGTCGTTACA	AAACCTGAAG	TAAAATTTGG	TGACTACAAA	AACCTTGAAG	TATCAGTTGA	660
	TGTAGAAAAA	GAAGTAAGTC	ACGCTGATGT	CGAAGAGCGT	ATCGAACCGG	AACGCAACAA	720
	CTGGGCTGAA	TTGGTTATCA	AGGAAGCTGC	TGCTGAAAAC	GGCGACACTG	TTGTGATCGA	780
30	CTTCGTGGT	TCTATCGACG	GTGTTGAATT	TGACGGTGGA	AAAGGTGAAA	ACTTCTCACT	840
	TGGACTTGGT	TCAGGTCAAT	TCATCCCTGG	TTTCGAAGAC	CAATTGGTAG	GTCACTCAGC	900
35	TGGCGAAACC	GTTGATGTTA	TCGTAACATT	CCCAGAAGAC	TACCAAGCAG	AAGACCTTGC	960
	TGGTAAGAAA	GCTAAATGTG	TGCAACATCT	CCACGAAGTA	AGACTCAAGG	AAAGTCCACG	1020
	TCTTGACGAT	GAACCTTGCA	AAGACATTTA	TGAAGAAGTT	GAACCACTTG	CTGCTTTGAA	1080
40	AGAAAAATAC	CGCAAGAAAT	TGGCTGCTGC	TAAAGAAGAA	ACTTACAAAG	ATGCAGTTGA	1140
	AGGTGCAGCA	ATTGATACAG	CTGTAGAAAA	CGCTGAAATC	GTAGAACTTC	CAGAAGAAAT	1200
45	GATCCATGAA	GAAGTTCACC	GTTCAGTAAA	TGAATTCCTT	GGGAACCTTG	AACGTCAAGG	1260
	GATCAACCTC	GACATGTACT	TCCAAATCAC	TGGAAC TACT	CAAGAAGACC	TTCAACAACA	1320
	ATACCAAGCA	GAAGCTGAGT	CAGTACTTAA	GACTTAACCTT	GTATCGAAG	CAGTTGCCAA	1380
50	AGCTGAAGGA	TTTGATGCTT	CAGAAGAAGA	AATACAAAAA	GAAGTTGAGC	AATTGGCAGC	1440
	AGACTACAAC	ATGGAAGITG	CACAAGTTCA	AAACTTGCTT	TCAGCTGACA	TGTTGAAAAA	1500
55	TGATATCACT	ATCAAAAAAG	CTGTTGAATT	GATCACAAGC	ACAGCAACAG	TAAATAAATC	1560
	TTAATAAGAA	GAAACCCAC	CTGAATTTGGT	GGGTTTTCTG	ATGCACATTT	TTCCAAAAAT	1620
	CTCTTTTGAAG	TCTGTTGCTG	TAATCCCAAT	CATGGCTGGG	ATGCGGTCCC	AGTTTCTCTC	1680
60	GTTTAGAGAT	TAGAGTTGTC	CAGAGGCAT	GTTTGTGACT	GTTTCAGAGA	CAGCTTTGTTG	1740

WO 98/26072

PCT/US97/22578

-104-

```

CTTTTCTTCA ACATTCTCCA GTAGATCACT GAAGCGTTCA ATCAGATAGG TTTTTCGGGC 1800
AGTTCCGATG TGTGGGGTAG CATAGTCGAA GGCTTSTAAT TCGCCTAGTA AGATGAGTTT 1860
5 GCTTTTGGCA CGTSTAATGG CTGTGTAGAT GAGATTTCGC TCCAGCATAC GTCGGCTAGC 1920
ACTAGTAATC GGTAGGATGA CAACTGGGAA CTCACTCCCG TGAGACTTAT GAATACTCAT 1980
10 GGCATAGGCC AAGCGAATCT TGTACCATTC GTTACGGGGG TAAGAGACTT CATTACTATC 2040
AAAATCAATG ACAATCTCGT CTTGTTTCGA TTCGGTGTAT TTACCGAGAA TCAGSTCTGT 2100
GATAGCTCCT AAATCCCAT TAAAGACATT GATTTCAGCA TCGTTAACCA AATGAATGAC 2160
15 CTTGTCTCTC TTACGATAGT GACACTGAGG AGCTTCAAAA CTGAGTTGAT CTTTTTGTGG 2220
GGGATTGAGC AGGTCTTGCA TGAGCTGATT GATAGCATCA ATCCCTGCCG TCCTTCGGTA 2280
20 CATAGGAGCC AGAAGTGGG TATCAGGGC GGGAAATACCA TTTCTGAGGG CGGCACCTAA 2340
GATTTTTTCA ATGTGGCGAG GAATATGGCC ACTAGCAATT TCAAAGTAGG AACGGTCAGC 2400
TTTTTTTTGG GTGAAATCAG CTGGCAAGAT GCCCTGTGCA ATCTGACTAG CTAGGGTGAC 2460
25 GATGGTTGAT TCTT 2474

```

(2) INFORMATION FOR SEQ ID NO:49:

```

30 (i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 716 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear
35 (ii) MOLECULE TYPE: DNA (genomic)
    (iii) HYPOTHETICAL: NO
40 (iv) ANTI-SENSE: NO

```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

```

ATCAAAATTA ACGTATTCTT TTTGAAGTTC AAGAACTTCT TCCATTGTTG AGCATTCTGT 60
AAGGGCAGGG TTTGCGTACT CTTCCATCTT AGCTGTGTCG AGTTTCTTCA TCAAGCTGCG 120
50 TGTACGAAGT ACAGATGTTG CTGACATAGA GAACTCATCC AAGCCCATTC CGACAAGAAG 180
TGGAACAGCT TGTGGTCCAC CAGCCATCTC ACCACACATA CCAGCCCAT TACCTTCAGC 240
55 GTGAGCTGCT TTGATCACAT TGTAAATCAA GCGTAGGATT GATGGGTTGT ATGGTTGGTA 300
AAGGTATGAA ACTTGTTCGT TCATACGGTC TGCTGCCATT GTATATTGGA TCAAGTCGTT 360
TGTACCAATT GAGAAGAAGT CAACTTCTTT AGCAAATTGG TCTGCAAGCA TAGCCGCTGC 420
60 AGGAATCTCG ATCATGATAC CAACTTGAAT GTTATCCGCA ACTGCAACAC CTTCAGCAAG 480

```

-105-

AAGGTTTGCT TTTTCTTCAT CAAAGACTGC TTTCGCTGCA CGGAATTCTT TCAAGAGCGC 540  
5 AACCATTTGGG AACATGATAC GCAATTGACC GTGAACAGAC GCACGAAGAA GAGCAGGGAT 600  
TTGTSTGCGG AACATAGCAT CTCAGTCTC AGAGATAGAG ATACGAAGAG CACGGAATCC 660  
AAGGAATGGG TCATTCTGTA GGCATATCGA AGTAAGGAAG TCCTTATCTC CACCGA 716

10 (2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 962 base pairs  
(B) TYPE: nucleic acid  
15 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

30 AGTAACCTAA ATCAATTATG GTGTTATGAG TCTTGGTGTG CCCAAAGTGC TGACGTAAC 60  
ATCTCAGCTG AAGGTGCAGA TGCAGATGGC CTATCGCTGC AATCTCAGAA ACAATGGAAA 120  
AAGAAGGATT GGCAATAGG AAATGACAGA AATGCTTAAA GGAATCGCAG CATCTGACGG 180  
35 TGTTGCAGTT GCAAAAGCAT ATCTACTCGT TCAGCCGGAT TTGTCAATTG AGACTATTAC 240  
AGTCGAAGAT ACAACGCAG AAGAGCTCG CCTTGATGCC GCTCTACAGG CATCAAGA 300  
40 CGAGCTTTCT GTTATTCGCG AGAAGCAGT AGGTACGCTC GGTGAAGAAG CAGCTCAAGT 360  
TTTTGATGCT CACTTAATGG TTCTTGCTGA CCCAGAAATG ATCAGCCAAA TCAAGGAAAC 420  
TATCCGTGCG AAGAAAGTGA ATGCAGAAGC AGGTCTGAAA GAAGTTACAG ATATGTTTAT 480  
45 CACTATCTTT GAAGGCATGG AAGACAACCC ATACATGCAA GAACGCGCAC GGATATCCGC 540  
GAGGTGACAA AAGTGTATT GGCAACCTT CTGGTAAAA AATTGCCAAA CCCAGCTTCT 600  
50 ATCAATGAAG AAGTATTGT GATTGCGCAT GACTTGACTC CTTGAGATAC AGCTCAATTG 660  
GACAAAAACT TTGTAAAAAG TTTGTAAAC AACATTGGTG GACGTACAAG CCATCTCAGT 720  
ATCATGGCAC GTACACTTGA AATTGCTGCT GTATTAGGTA CAAACAACAT CACTGAAATC 780  
55 GTTAAAGAGC GTGACATCCT TGCTGTTAAC GGGATCACTG GAGAAGTGAT TATCAACCCA 840  
ACAGATGAAC AAGCGGCAGA ATTTAAAGCA GCTGGTGAAG CCTATGCGAA CAAAAGCTG 900  
60 AATGGGCACT TTTGAAAGAT GCTCAACAGT GACTGTGAC GGTAACACTC GAGTTGGCTG 960  
CC 962

-106-

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2702 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

20	GATCGTTTCC GTGGCTTGAT CGGAAGCATG TTTGACGAAT AAAGAGGAAA AATAAATTAT	60
	GACATTTTCA TTTGATACAG CTGCTGCTCA AGGGGCAGTG ATTAAGTAA TTGGTGTGCG	120
25	TGGAGGTGGT GGCAATGCCA TCAACCGTAT GGTGACGAA GGTGTTACAG GCGTAGAATT	180
	TATCGCAGCA AACACAGATG TACAAGCATT GAGTAGTACA AAAGCTGAGA CTGTTATTCA	240
	GTTGGGACCT AAATTGACTC GTGGTTTGGG TGCAGGAGGT CAACCTGAGG TTGGTCGTAA	300
30	AGCCCGTGAA GAAAGCGAAG AAACACTGAC GGAAGCTATT AGTGGTGCCG ATATGGTCTT	360
	CATCACTGCT GGTATGGGAG GAGGCTCTGG AACTGGAGCT GCTCCTGTTA TTGCTCGTAT	420
35	CGCCAAAGAT TTAGGTGCGC TTACAGTTGG TGTGTAACA CGTCCCTTTG GTTTTGAAGG	480
	AAGTAAGCGT GGCAATTG CTGTAGAAG AATCAATCAA CTCCTGAGC ATGTAGACAC	540
40	TCTATTGATT ATCTCAACA ACAATTGCTT TGAAATTGTT GATAAGAAAA CACCGCTTTT	600
	GGAGGCTCTT AGCGAAGCGG ATAACGTCTT TCGTCAAGGT GTTCAAGGGA TTACCGATTT	660
	GATTACCAAT CCAGGATTGA TTAACCTTGA CTTTGCCGAT GTGAAAACGG TAATGGCAAA	720
45	CAAGGGAAT GCTCTTATGG GTATTGGTAT CGGTAGTGA GAAGAACGTG TGGTAGAAGC	780
	GGCAGCTAAG GCAATCTATT CACCCTCTCT TGAAACAAT ATTGACGGTG CTGAGGATGT	840
50	TATCGTCAAC GTTACTGGTG GTCTTGACTT AACCTTGATT GAGGCAGAAG AGGCCTTACA	900
	AATGTGGAAC CAGGCAGCAG GTCAAGAGT GAACATCTGG CTCGGTACTT CAATTGATGA	960
	AAGTAGCGT GATGAAATTG GTGTAACAGT TGTGCAACG GGTGTTGCTC AAGACCGCGT	1020
55	AGAAAGGTT GTGGCTCCAC AAGCTAGATC ACCGCGCCTA GGATAACAA TTTAGCAATC	1080
	AAGATAAACC AAAACATCAT AACAACAAGA AGAACGGAAC CTAAATTCG GACATCCACC	1140
	AAATGATGGA CATAGTAATT GAGATAACTA GAGAACAGAG TTAGTACACC TAAATCACC	1200
60	AAGAGAACAA AGGCACTGCC TGGTAGGGTA TAGCTAATTT TCCTGTTAGA TAGATTGGGA	1260



WO 98/26072

PCT/US97/22578

-107-

AGAAAATAAT AAAGCATGAC CAAGATAGCA AAGAGGAGGG CGTAAATCAG AGGACCTGCC 1320  
 AACCCCTGTA AAGCCTGATA GATAATGCCA TCTTTTGTC AATAATGAGC AAGTAAAGCC 1380  
 5 AAAATCATCT GACCAATAAA GATCAAAAAC AAGGCAACG CAAAGAGGAA CTGCAAGCCA 1440  
 AAAGTACTA GAGAGCTTAG CATCTGATGG GAAATAAGTC CAGGACTCTT TTCGACGCCA 1500  
 10 TAAGCCTTGT TAAAGGCTTT TTGCAAGAAA TTTATAGATT TTGAAAACT CCATAACGCC 1560  
 GATAAAACAG AAAAATCTAA TAAACCTGTT GAAGGTTGCG TCAAGAGCTT CTCTGGCTAT 1620  
 15 TTTTCCACA CCTTCATAGA GGCTTGCGGG CAGGACGCTT TTCATAAAGC CCAGAAATTC 1680  
 TCCACAGGA ATCTGAAAT AGGGGAGGAT ATTGACCACC ACCAAAAGCA GGGGGAATAT 1740  
 CGAAATCAAC CANTAGTAGC CTACTGCGAC ACTGGTCAA CTCACTATCT GATGCTTGAT 1800  
 20 AATAATGCAA AAAAGCTTTT AATAAAGGCT TGTCTATCAG CTCTTTCAC CACTTTTCA 1860  
 TGTCACTACT CTTCATTAT AATCTTATAC TCAATGAAA TCAAGAGCA AACTAGAAAG 1920  
 CTAGCCGCAA GCTGCTCAA AACTGTTTT GAGGTTGTAG ATAAGACTGA CGAAGTCAGT 1980  
 25 CACATACATA CGGTAAAGCG ACGCTGACGT GGTTTGAAGA GATTTTCGAA GAGTATTAC 2040  
 TAATTTCTCT TTACCAATTC CACCATATCA TACGGTAGGG TATTGGCAGC TTCCTTCAAG 2100  
 30 GAATAGTTCT CTAAGTTATT TACATTTTGT CGTAATTTCT TGGCATACTT AGTTGTAATT 2160  
 AATCGTTTTT CTTCGTATTC GAAATCAAC TTGCGCTCCA GATAATAGCC TCTCAGCATT 2220  
 35 TCATTGATAT TGTGGGTTT GACACGATTG ATAACCCGTT CGACAAAGGC ACCACTGCTG 2280  
 ATAATAGTTG TTTCTCGAAG ACGAGACTCC TGCATAAAAC TAATCAAAGA GCGCTGTGAG 2340  
 ACTCCCTTCA GGTTTTCCAA ACTTTCAATA ATCATCTCCG TATTGGCAAG ATAGAGCTCT 2400  
 40 GCAATTTGGT CATAATCAAG AGCAGGAGA GGGCTTTGCT CTTGTCCCTT CCAGCTACGG 2460  
 AAGGTCCTTC CAAGAGTAAA AACTTCATGA AGGAGAAAAA GTAAAATCCT CAAGGAAACA 2520  
 AGAAAATAAT AGTCACTCT TGAGGCAAGT TTACGATGTA TTCCTTGTC TATATTTTTC 2580  
 45 AGATAACGTT GGTAACTCG GTAAGCACGA TTGCTAATGT TCCCTCTTC ATAGGCCTGT 2640  
 TCCAAACCAT CACTTTCAA? ACTAAGAATC AAGAGTTTCA AAGCAGCCCA GTCTTCTTGA 2700  
 50 TC 2702

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6217 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 55  
 60 (ii) MOLECULE TYPE: DNA (genomic)

-106-

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

10	GAATCCAAG AAGCTAGCCA AGAAAGTCGT GAACGTAGTG ATCCGCTAAA TAGTTATCTC	60
	CTTTTGTACG GCTCCTTGAC GAAAGAAAAG CTTGCCGATA AATTAGGAGA TTTGGGTTAT	120
15	AAAGCAAGTG CTGACGAAA GATACCGCCC TATTTCTTGT CTTTCGAAT ATTACTAAAT	180
	CCCCATTATT TAATTAGTTT AGCAATATTT GGCTTATCTT TCTTTGCTTT AGTGATTATC	240
	ACTCGGATTA AGGAAATGAG AGCAGCAGGT ATAAACCTCT TTTCTGGTCA GACTCTCTTA	300
20	TCCATCATGG GGCATTCTTT ATCTACTGAT ATCAAAATGGC TCCTTCTATC AGCCCTCCTT	360
	TCCTTCCTAG GTGGGGGTGT CGTTCTTTTT AGTCAAGGTT TGTTTTATCC TATCTTGTTA	420
	GCCACCTATG GTTTTGGGAT TAGTTTCTAT CTGTTGTTTT TATTGGCGAT TTCAAATTTA	480
25	CTAATGCTTC TTTATCTAAT GAGTTTGAAT ACAAGCATT AGTTCGCCGT ATTAGGGGGA	540
	GATTCGCCCT GAACCTGAT GAATAACCCA TTGTTTCAGT AGACCTGTTT TTTCAGTAGG	600
30	ATACGCTTTA AGACAGGTTG ACGTCTTACC AACGATTGAA AGAACTTGAA ATTTCAGACA	660
	AGATGGCAGG ATAGAGTAGA CTATTATCAC GATTTCCTTT GACTTAGGTT ATAGAGGTTG	720
35	AGATTCAGAA AATCAGAGCA AGTGGTATGC CTTTACCAGG GGAGCAGCGA AGAAGAACA	780
	GCTCTTTATG TAAAGGATAA TCTGCTCCAT TTTGCCAATC CACAAGGAAA AAATGAACAG	840
	GGAGAGACAC TGGATACCTA TAGTCCAGAT GCTAATACGC TCTATGTTAG TCCAGTTTAT	900
40	TTGGACAAGG AAAAGGTCGT GGTAGATGCT GAGACCAAAC AGAAGTTAGC CCATCTCCAA	960
	AAAGGTGAGT TTATCCTCTT GCTCCAGAA CATTGCGCT CTGAGAAGC AGAACTTAAG	1020
	AAAGTTTTTG AAGAAAGATT GAGTTATTAT GGAAATCTG GTGAGGAGGC AAGTGCTCCT	1080
45	TTGGATTATG AGATGAAAGC GCACGTTAGT TATCTTTCAA TGGGAGAAAA GCGGTTTGTT	1140
	TATAATAACG GTGGAATCC CGTATCTACT CAGTATTGTA CTGATCCGAT TTTAGTTGTA	1200
50	TTCAAGCCCA CTTCTACAGG TGATAGTTTT ATATCCCTAT CTAGTTGGTC TATCAATGCT	1260
	GGAAACAAC TCTTTATCAA AGGATATGAG AGTGGGCTAG AACTCTTGAA GAAAGCTGGA	1320
55	ATTTATGAGC AAGTATCCTA TCTTAAAGAA GGAAGAAGTC TTTATCTAAT TCGTTATAAT	1380
	GAAGTTCAAA CTGAACAGC AACTTTAATC TTAGGAGCTA TTGTTGGGAT AGCTAGTTCC	1440
	TTGTACTCTT TTTATCTGT CAATCTTCTA TATTCGAGC AATTCGCCG AGATATCTTG	1500
60	ATTAAACGAA TTTCAGGTTT ACGATTTTTT GAAACACATG CTCAGTATAT GGTTAGTCAA	1560

-109-

	TTTGCCAGTT	TTGCTATTGG	TGCTAGTCTC	TTTATTTTAA	GCAGTCGAGA	CTTGGTGATT	1620
	GGCTTGCTCA	CTTTATTAGT	CTTCTAGCT	AGTGCAGTTT	TGACGCTTTA	CCGTCAAAGCG	1680
5	CAGAAAGAA	CTCGTGTTTC	TATGACAATT	ATGAAAGGAA	AATAGGATGA	TTGAACATAA	1740
	GAATATATCT	AAAAAATGG	AAGCCGTCAG	CTATTTCAG	ATACGAATCT	TCATTTTGAA	1800
10	GGTGGGAAA	TTTATGCCCT	AATCGGTACA	AGTGGCTGTG	GTAAGACAAC	ACTCTCGAAT	1860
	ATGATTGAGC	GATTGGCGCC	ATATGACAAA	GGGCAATCA	TCTATGATGG	CACCTCTCTT	1920
	AAGGACATCA	AGCCCTCTGT	TTTCTTTAGA	GATTACTTAG	GATACTTATT	TCAAGATTTT	1980
15	GGCTTAATTG	AAAGCCAAAC	CGTCAAAGAG	AATCTCAATC	TGGGTTTAGT	TGGTAAWAAAG	2040
	TTGAAGGAAA	AAGAGAAAA	CTCTTTGATG	AAACAAGCTC	TAAACCGTGT	AAACCTCTCT	2100
20	TATTTGGATT	TGAAGCAACC	TATATTTGAG	TTATCAGGAG	GAGAAGCACA	ACGTGTTGCA	2160
	CTAGCGAAGA	TAATTTTAAA	GGATCCGCGT	TTGATTTCTG	CAGATGAACC	AACCGCTTCC	2220
	TTAGACCCCA	AAAAATCTGA	GGAAATTACT	TCCATCCTAG	AATCTTTAAA	AAATCCGAAT	2280
25	CGGACCATTA	TTATTGCGAC	CCACAATCCT	CTGATTTGGG	AGCAAGTGGG	TCAGGTCATT	2340
	CGAGTTACCG	ATTATACACA	TAGATGATAT	GGTAAGATTG	AGTTAGAAGA	AAGAGTCACA	2400
	AACACACTTT	GTGGCTTTTT	TATTTCCATA	AAAATGGTAA	AATAGTAGGA	GTAGAAATGA	2460
30	GTTCGAGACA	TGAAAGTAAT	AGATCAATTT	AAAAATAAGA	AAGTTCTTGT	TTTAGGTTTG	2520
	GCCAAGTCTG	GTGAATCTGC	AGCTCGTTTG	TTGGACAAGC	TAGGTGCCAT	TGTGACAGTA	2580
35	AATGATGGGA	AACCTTTTCA	GGACAATCCA	GCTGCCCAAA	GTITGCTGGA	AGAAGGGATC	2640
	AAGGTCATTG	CAGGTGGCCA	TCCTTTGGAA	CTCTGGATG	AAGAGTTTGC	CCTTATGGTG	2700
40	AAAAATCCAG	GTATCCCTCA	CAACAATCCC	ATGATTGAAA	AGGCTTTGGC	CAAGAGAATT	2760
	CCAGTCTTGA	CTGAGGTGGA	ATTGGCTTAT	TTGATTTTCA	AAGCACCAGT	TATTGATATC	2820
	ACAGGATCGA	ACGGTAAGAC	AACCACAACG	ACTATGATTG	GGGAAGTTTT	GACTGTCTGT	2880
45	GGGCAACATG	GTCTTTTATC	AGGGAAATAT	GGCTATCCTG	CCAGTCAGGT	TGCTCAAATA	2940
	GCATCAGATA	AGGACAGGCT	TGTTATGGAA	CTTTCTTCTT	TCCAATCAT	GGGTGTTCAA	3000
	GAATTCATC	CAGAGATTGC	GTTTATTACC	AACCTCATGC	CAACTCATAT	CGACTACCAT	3060
50	GGGTCAATTT	CTGAATATGT	AGCAGCGAAG	TGGAATATCC	AGAACAAGAT	GACAGCAGCT	3120
	GATTTCTCTG	TCTTGAACTT	TAATCAAGAC	TTGGCAAAAG	ACTTGACTTC	CAAGACAGAA	3180
55	GCCACTGTTG	TACCATTTTC	AACACTTGAA	AAGGTTGATG	GAGCTTATCT	GGAAGATGGT	3240
	CAACTCTACT	TCCGTGGTGA	AGTAGTCATG	GCAGCGAATG	AAATCGGTGT	TCCAGGTAGC	3300
	CACAATGTGG	AAAATGCCCT	TGCGACTATT	GCTGTAGCCA	AGCTTCGTGA	TGTGGACAAT	3360
60	CAAAACATCA	AGGAAACTCT	TTCAAGCCTTC	GGTGGGTGCA	AACACGCTCT	CCAGTTTGTG	3420

WO 98/26072

PCT/US97/22578

-110-

	GATGACATCA AGGGTGTAA ATTCTATAAC GACAGTAAAT CAACTAATAT CTGGGCTACT	3480
5	CAAAAAGCCT TATCAGGATT TGACAACAGC AAGGTCGTCT TGATTGCAGG TGGTTTGGAC	3540
	CGTGGCAATG AGTTT'GACGA ATTGGTGCCA GACATTACTG GACTCAAGAA GATGGTCATC	3600
	CTGGGTCAAT CTCGAGAAGC TGTCAAACGG GCAGCAGACA AGGCTGGTGT CGCTTATGTG	3660
10	GAGGCGACAG ATATTGCAGA TCGGACCCGC AAGGCCTATG AGCTTGCAGC TCAAGGAGAT	3720
	GTGGTCTTTC TTATCTCTGC CAATGCCAGC TGGGATATGT ATGCTAAGTT TGAAGTACGT	3780
	GGCGACCTCT TTATCGACAC AGTAGCGGAG TTAAGAAGAT AAAATATGAA AAAAATTGTC	3840
15	TTTACAGGTG GGGGGACGGT TGGACACGTT ACCCTCAATC TTTTGTAAAT GCCCAAGTTC	3900
	ATCGAAGATG GTTGGGAAGT CCACATATAT GGGGACAAGC GTGGTATCGA ACACCAAGAA	3960
20	ATCCTTAAGT CAGGTTTGA TGTCACTTTC CACTCCATTG CGACTGGGAA ATTGCGTCGC	4020
	TATTTCTCTT GGCAAAATAT GCTGGACGTC TTCAAAGTTG GCTGGGGAAT CGTCCAATCG	4080
	CTCTTTATCA TGTTGCGACT TCGTCCACAG ACCCTTTTTT CAAAGGGGGG CTTTGTCTCA	4140
25	GTACCGCCTG TTATCGCAGC GCGTGTGTCA GGAGTGCCTG TCTTTATTC ACGAATCTGAC	4200
	CTGTCTATGG GCTTGGCCAA TAAATCGCC TATAAATTG CCACTAAGAT GTATTCAACC	4260
30	TTTGAGCAAG CTTTGAGTTT GTCTAAGGTT GAGCATGTGG GAGCAGTGAC CAAGGTTTCA	4320
	GATCAAAAAA ATCCAGAACC AGATGAATTG GTGGATATTC AAACCACTT TAATCATAAA	4380
35	TTGCGGACTG TATTGTTTGT TGGCGGTTCT GCAGGTGCTC GTGTCTTTAA CCAATTGGTG	4440
	ACAGACGATA AGAAAGAACT AACAGAGCGC TACAATATTA TCAATCTAAC TGGAGATTCT	4500
	AGTCTGAACG AGTTGAGCCA AAATCTTTT CGTGTGACT ATGTGACCGA TCTCTATCAA	4560
40	CCCTTGATGG AATTGGCTGA TATTGTTGTG ACACGAGGTG GTGCCAATAC GATTTTGGAG	4620
	CTCTTGGCGA TAGCAAAATT GCATGTCAAT GTGCCGCTTG GTGCTGAAGC TAGTCTGGT	4680
	GACCAGATTG AAATGACAGC TTACTTTGTA AAAAAAGGCT ATGCAGAAGA CCTTCAAGAA	4740
45	AGCGATTTGA CCTTGGATAG TTTGGAAGAG AAGCTTTCTC ACTTACTAAG TCACAAGGAA	4800
	GATTACCAAG CTAAGATGAA AGCTTCTAAG GAATTGAAAT CTCTAGCAGA TTTTATCAA	4860
50	TTGTTGAAAA AAGATTATC ATAAGGAAAG TAAATGTCAA AAGATAAGAA AATAGAGGAC	4920
	AAAGAAACCC TCGAAGAAAT GAAAGAGTTA TCAGATGGC AGAAACGAAA CCAAGAAATAT	4980
55	CTAAAAAAGA AGGCTGAAGA AGAGTGGCT CTAGCTGAGG AGAAGGAAAA GGAAGACAA	5040
	GCTCGAATGG GAGAAGAATC TGAGAAGTCA GAGGACAAAC AGGACCAGGA GAGTGAAACA	5100
	GACCAGGAAG ATTCAGAATC AGCTAAGGAA GAGTCTGAAG AAAAAGTAGC ATCCTCAGAG	5160
60	GCTGACAAGG AGAAAGAAGA ACCAGAGTCT AAAGAGAAGG AGGAACAGGA TAAAAAGCTT	5220

WO 98/26072

PCT/US97/22578

-111-

GCTAAAAAGG CTACAAAGGA AAAACCAGCC AAAGCAAGA TTCCTGGTAT CCATATCTTG 5280  
CGAGCCCTCA CGATTTTATT TCCAAGCTG CTTTATTGA TTGCTCTGC CTACTTGCTC 5340  
5 AGTCCTTATG CGACCATGAA AGATATTCGT GTTGAGGGA CGGTGCAAC TACAGCTGAT 5400  
GATATTTCGAC AGGCTTCAGG CATTAGGAT TCGGATTATA CGATTAACT TCTGCTAGAC 5460  
AAGGCAAAAT ATGAAAAGCA GATTAAGTCT AACTATTGGG TTGAATCAGC TCAACTTGTC 5520  
10 TATCAATTTC CAACTAAGTT CACTATTAA GTCGAAGAA ATGATATTGT GGCCTACTAT 5580  
ATTCIGGTG AAAATCATTA TCCTATTCTT TCCAGTGGT AGCTTGAGAC TAGTCTGTG 5640  
15 AGTCTGAACA GTTTACCAGA AACTTATTTA TCAGTTCTCT TTAATGATAG TGAACAAATC 5700  
AAGGTTTTG TCTCAGAACT TGCTCAAAT AGCCAGAAC TCAAGCGGC TATCCAAAAG 5760  
TGGAATTAG CCCCAAGCAA GGTGACATCC GATTTAATC GATTGACCAT GAATGATTG 5820  
20 GACGAAGTCT TGTTCTCTCT ATCTGAAATG AGTAAGAAAC TGCCATATTA CAGTAAGATT 5880  
AAGCCAAT TGTCAGAAC GAGTGTGGT GACATGGAAG CTGGAATTTA CAGTTACACT 5940  
25 GTGGCGGATA AATTAATTAT GGAGGCTGAG GAAAAAGCCA AACAGAGGC CAAGGAAGCT 6000  
GAGAAAAAC AAGAAGAAG ACAGAAAAA CAAGGAAGG AGAGCAATCG AAATCAACA 6060  
AATCAGGCTT CATCGCTCG CTAGGTTTAC CTTTCTCTT ATAGTTCTT AGTGACCATG 6120  
30 TTTTACGTT TAATATTGA CATTGTTTT TCTTATGTT ACATCTGCAA TGTAACTCAT 6180  
TACAAAATA TTTTGTATGA AGAAGTAAC ACATATG 6217

35 (2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1491 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CTTGACACTT GATTGCGACT GTTGAACTT ATCTCTCAA GAAAAACAG TGAAGATGT 60  
55 GAGTCTGCTG TCAGCAAGCT TGAAGTAGC ACATCTGAG AACATTGGAT CCATCTGCAG 120  
TTTCTCGTGG GTCTAGCTTG GATCGTGAT ACAATGGTCT TTTGACTCTT GCTGGCGGTA 180  
60 AAATCAGAGA CTACCGTAAG ATGGGTGACG AGCGCGTAT GAGCGCGTGG TTGACATCCT 240

-112-

CAAAAGCAGAA TTTGACCGTA GCTTTAAATT GATCAATTCT AAAACTTACC CTGTTTCAGG 300  
TGGAGAAATTG AACCCAGCAA ATGTGGATTG AGAAATCGAA GCCTTTGCGC AACTTGGAGT 360  
5 TTACAGTGGT TTGGATAGCA AGGAAGCTCA TTACCTAGCA AATCTTTACG GTTCAAATGC 420  
ACCGAAGGTC TTTCGACTTG CTCACAGCTT GGAACAAGCG CCAGGACTCA GCTTGGCAGA 480  
TACTTTGTCC CTTGACTATG CAATGCGCAA CGAGTTGGCT CTTAGCCGAG TTGACTTCCT 540  
10 TCTTCGCTGT ACCAACCATTA TGCTCTTAT GCCTGATAGC TTGGATAGCA TCGTTGAGCC 600  
AGTTTGGAT GAAATGGGAC GATTCTATGA CTGGACAGAA GAAGAAAAAG CAACTTACCG 660  
15 TGCTGATGTC GAAGCAGCTC TCGCTAACAA CGATTTAGCA GAATTAAGAAA ATTAAGAAAA 720  
AATAAAGAG GTGGAGGGCA GCATTCCTTG TCGCCCGTCC CTTCTTTTAA ATGGAGACAG 780  
AAAGATGATG AATGAATTAT TTGGAGAATT TCTAGGGACT TTAATCCTGA TTCTTCTAGG 840  
20 AATGGTGTG GTTGCAAGTG TGGTCTTCC TAAACCAAG AGCAATAGCT CAGGTGGAT 900  
TGTGATTAAT ATGGGTGGG GGATTGCAGT TGCGGTTGCA GTCTTTGTAT CTGGCAAGCT 960  
25 CAGTCCAGCT CATTAAACC CAGCTGTGAC CATCGGTGTG GCCTTAAAAG GTGGTTTGC 1020  
TTGGGCTTCC GTTTTGCCTT ATATCTTAGC CCAGTTGCGA GGGGCCATGC TGGGTCAGAT 1080  
TTTGGTTTGG TTGCAATTCA AACCTACTA TGAGGCAGAA GAAATGCGA GCAATATCCT 1140  
30 GGCAACCTTC AGTACTGGAC CAGCCATCAA GGATACTGTA TCAAACCTGA TTAGCGAAAT 1200  
CCTTGGAAAC TTGTTTTGG TGTGACAAAT CTTTGCTTTG GGTCTTTACG ATTTTCAGGC 1260  
35 AGGTATCGGA ACCTTTGCAG TGGGAACCTT GATTGTGGGT ATCGGTCTAT CACTAGTGG 1320  
GACAACAGGT TATGCCTTGA ACCCAGCTCG TGACCTTGGG CCTCGTATCA TGCACAGCAT 1380  
CTTGCCAAAT CCAARCAAGG GAGACGGAGA CTGGCTTAC GCTTGGATTG CTGTTGTAGG 1440  
40 CCCTGTTATC GGAGCAGCCT TGGCCGTGCT TGTATTGTCA CTTTCTAAT C 1491

(2) INFORMATION FOR SEQ ID NO:54:

- 45 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1229 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- 50 (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- 55 (iv) ANTI-SENSE: NO
- 60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

-113-

5    ACAAACGATA ATGTCATCGA TCTCTTTGAA CACATCTTTA AGGAATGTTT AACGAAAACA    60  
      TTGTGATGGC GGGCAAGGTC AATCTCTTGA ATTTTGCCAA TCTAGCAGCC TATCAGTTCT    120  
 10    TTGACCAACC GCAAAAGGTG GCCTTGGAGA TTCGTGAGGG GTTGCGTGAG GATCAGATGC    180  
      AAAATGTTTC GTTTCAGAGC GGTCAAGAGT CCTGTTTAGC TGACCTAGCG GTGATTAGTA    240  
 15    GTAAGTTCCT CATTCTTTAT CGGGGAGTTG GAATTTAGC CATTATCGGT CCAGTTAATC    300  
      TGGATTACCA ACAGCTAATC AATCAAATCA ATGTGGTCAA CCGTGTTTTG ACCATGAAAT    360  
      TGACAGATTT TTACCGCTAC CTCAGCAGTA ATCATTACGA AGTACATTAA GATTGAAATC    420  
 20    ATTAAGAGAG GCGAACATGG CCCAAGATAT AAAAAATGAA GAAGTAGAAG AAGTTCAGAA    480  
      AGAGGAAGTT GTGGAACAG CTGAAGAAAC AACTCCTGAA AAGTCTGAGT TGGACTTGGC    540  
      AAATGAACGT GCAGATGAGT TCGAAAACAA ATATCTTCGC GCTCATGCAG AAATGCAAAA    600  
 25    TATCCAACGC CGTGCCAATG AAGAACGTCA AAACCTTGCAA CGTTATCGTA GCCAGGACTT    660  
      GGCAAAAGCA ATCTTACCAT CTCTTGACAA CCTTGAGCGT GCACCTGACG TTGAAGGTTT    720  
      GACAGATGAT GTGAAGAAGG GCTTGGCGAT GGTGCAAGAA AGCTTGATTC ACGCTTTGAA    780  
      AGAAGAAGGA ATTGAAGAAA TCGCAGCAGA TGGCGAATTT GACCATAACT ACCATATGGC    840  
 30    CATCCAAACT CTCCAGGAG ACGATGAACA CCGAGTAGAT ACCATCGCCC AAGTCTTTCA    900  
      AAAAGGCTAC AAACCTCCATG ACCGCATCCT ACGCCAGCA ATGGTAGTGG TGATATAACTA    960  
      AGATACAAAC GCTCGTAAAA AGCTCGCAGT AAAAATAGGA GATTGACGAG TGTTCGATGA    1020  
 35    ACACAAGAAA ATCTATCTTT TTTACTCAGA GCTTAGGGCG TGTTGCGATTC GGCAATCTCG    1080  
      ACGGTAGCTA AAGCAACTCG TCAGAAAACG GCAATCGCTA TGACGTTTGC CTAGCTTCCT    1140  
      TACTAACTCG TCGTCGAAAT AAAATCGATT TCGACTCCTC GTGTCGCAAT TTACATAATA    1200  
 40    GAAACTTGT CCGAACGACA TAAACTATG    1229

(2) INFORMATION FOR SEQ ID NO:55:

45    (i) SEQUENCE CHARACTERISTICS:  
         (A) LENGTH: 5816 base pairs  
         (B) TYPE: nucleic acid  
         (C) STRANDEDNESS: single  
 50       (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

55    (iv) ANTI-SENSE: NO

60    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

WO 98/26072

PCT/US97/22578

-114-

	AAGAAGAAGA	CTGTATGGAT	AATCGACCAA	TTGGTTTTTT	GGATTGCGGT	GTCGGGGGCT	60
	TGACCGTTGT	GCGCGAGCTC	ATGCGCCAGC	TTCCCATGA	AGAAATCGTC	TATATTGGAG	120
5	ATTGCGCGCG	GGCGCCCTAT	GGCCCCCGTC	CTGCTGAGCA	AATTCGTGAA	TATACTTGCG	180
	AGCTGGTCAA	CTTTCTCTTG	ACCAAGSATG	TCAAAATGAT	TGTCATTGCT	TGTAACACTG	240
	CGACTGCGGT	TGCTGGGAA	GAAATCAAGG	CTCAACTAGA	TATTCCTGTC	TTGGGTGTAA	300
10	TTTTGCCAGG	AGCTTCGGCA	CCCATCAAGT	CCAGTCAAGG	TGGGAAATC	GGAGTATTG	360
	GAACGCCCAT	GACGGTACAA	TCAGACATAT	ACCGTCAGAA	AATCCATGAT	CTGGATCCCG	420
15	ACTTACAGST	GGAGAGCTTG	GCCTGTCCCA	AGTTTGCTCC	CTTGGTTGAG	TCAGGTGCCC	480
	TGTCAACCAG	TGTTACCAAG	AAGGTGGTCT	ATGAAACCTT	CGCTCCCTTG	GTTGGAAAGG	540
	TGGATAGCCT	GATTTTGGGC	TGTAATCATT	ATCCACTCCT	TCGCCCTATT	ATCCAAAATG	600
20	TGATGGGGCC	AAAGGTTTCA	CTCATCGATA	GTGGGGCAGA	GTGCGTACGG	GATATTTTCA	660
	TCCTATCAA	TTATTTTGAA	ATCAATCGTG	GTGCGGATCG	TGGACCACTC	CATCACCGTT	720
25	TTTACACAAC	AGCCAGTAGC	CAAAGTTTTG	CACAAATTGG	TGAAGAATGG	CTGGAAAAAG	780
	AGATTCATGT	GGAGCATGTA	GAATTATGAC	AAATAAAATT	TATGAATATA	AGGATGACCA	840
	GAAGTGGTAT	GTTGGGTCTT	ATAGTATTTT	TGGTGGCGTT	AACAGTTTGA	GCGACTATAA	900
30	GGCAGATTTT	CCTCTGTTTG	AATTCTCCAA	AATATTTTGA	GATGAAGAGT	TTGGTTTCCC	960
	GCTTTCAGTT	ACTGTTTTAC	GCTATGGTTC	TACCTACCGT	TTGTTCTCCT	TTGTGGTAGA	1020
35	CATGCTTAAT	CAAGAAATGG	GACGAAACTT	GGAAGTTATT	CAACGTCATG	GGGCCCTGCT	1080
	CTTGGTTGAA	AATGGGCAAC	TCTTGTATGT	AGAATTGCCT	AAAGAAGGGG	TCAATGTTCA	1140
	TGATTTCTTT	GAGACAAGCA	AGGTGAGAGA	AACCTTGTGG	ATTGCGACTC	GTAACGAAGG	1200
40	TAAACCAAG	GAATTCGGAG	CTATCTTTGA	TAAGTTAGGC	TACGATGTGG	AAAATCTTAA	1260
	TGACTACCTT	GACCTGCGTG	AAGTAGCAGA	AACAGGTATG	ACCTTTGAAG	AAAATGCCCG	1320
45	CCTTAAGGCA	GAAACCATTT	CTCAATTAA	GGGCAAGATG	GTTTTGGCAG	ATGATTCTGG	1380
	TCTCAAAGTC	GATGTCCCTG	GTGGCTTACC	AGGCGTCTGG	TCAGCTCGTT	TCGCAAGGTG	1440
	GGGAGCAACT	GACCGTGAAA	ATAATGCCAA	ACTCTTGCAC	GAATTGGCCA	TGGTCTTTGA	1500
50	ACTCAAGGAC	CGCTCGGCTC	AGTTCCACAC	AACCTTAGTC	GTAGCCAGCC	CAAAATAAGGA	1560
	AAGTTTAGTT	GTTGAAGCAG	ACTGGTCAGG	TTATATTAA	TTTGAACCTA	AGGGTGAAAA	1620
55	TGGCTTTGGC	TATGATCCCC	TCTTCTTGT	AGGAGAGACA	GGTGAGTCA	CAGCTGAATT	1680
	AACCCCTGGAA	AAAAAAAATA	GTCAATCTCA	CCGTGCCTTA	GCCGTTAAGA	AACTTTTGGA	1740
	GGTATTTCCT	TCATGGGCAA	GCAAACCATC	ATTGTAATGA	GCGATTCCCA	TGGCGATAGC	1800
60	TTGATTGTGG	AAGAAGTCCG	TGATCGCTAT	GTGGGCAAG	TCGATGCCGT	TTTTCTATAAC	1860



WO 98/26072

PCT/US97/22578

-115-

	GGCGATTCTG AACTACGTCC GGATTCTCCA CTTTGGGAGG GCATCCGCGT TGTAAAGGG	1920
5	AACATGGACT TCTACGCCGG CTACCCAGAA CGTCTGGTGA CTGAGCTTGG TTCGACCAAG	1980
	ATTATCCAAA CTGATGGTCA CTGTCTTGAC ATCAATTTC ACTTTCAAAA GTTGGACTAC	2040
	TGGGCTCAGG AGGAAGAGGC CGCTATCTGC CTCTATGGTC ACTTGCACTG GCCAAGTGCT	2100
10	TGGATGGAAG GCAAGATCCT CTTTCTAAAT CCAGGTTCTA TCAGTCAACC ACGAGGTACC	2160
	ATCAGAGAAT GTCTCTATGC TCGTCTGGAG ATTGATGATA GTTACTTCAA AGTGGACTTT	2220
15	TTGACACGAG ATCAGCAGGT GTATCCAGGT TTGTCCAAGG AGTTAGCCCG ATGATTGCCA	2280
	AGGAGTTTGA GACTTTCTTG TTGGGGCAGG AGGAAACTTT TTTGACCCCT GCTAAAAATC	2340
	TAGCTGTGTT GATTGATACC CACAATGCGG ATCATGCGAG CCTCTTGCTC AGTCAGATGA	2400
20	CCTATACCCG TGTTCCTGTT GTGACAGATG AAAAAACAGT TGTGGGACG ATTGGACTCA	2460
	GAGATATTAT GGCTTATCAG ATGGAGCATG ACTTGAGCCA AGAAATCATG CGCGATACGG	2520
25	ATATCGTTCA TATGACAAAA ACGGACGTAG CGGTTGTTTC GCCIGATTC ACCATTACGG	2580
	AGGTCTTGCA CAAGCTAGTA GATGAGTCTT TCTTACCGGT CGTGGATGCA GAGGGTATTT	2640
	TCCAAGGGAT TATTACGCGC AAGTCCATCC TCAAGGCCCT TAATGCCCTC TTGCATGACT	2700
30	TTAGTAAGGA ATATGAGATT CGATGCCAAT GAGAGACAGG ATTTGAGCCT TTTTAGAGGA	2760
	AAAGCAGGGC TTGTCTGTCA ATTCCAAGCA GTCCTATAAG TATGATTTGG AGCAATTTTT	2820
35	AGACATGGA GGTGACGCGA TTTCTGAGAC CAGTCTCAAG ATTTACCAAG CCCAGCTAGC	2880
	CAATCTAAAA ATCAGCGCCC AGAAGCGAAA GATTTGCGCC TGTAAACCAAT TTCTATACTT	2940
	TCTCTATCAA AAAGGAGAGG TGGACAGCTT TTACCGCTTG GAATTAGCTA AACAAAGCTGA	3000
40	AAAGAAGACG GAAAGGCCAG AGATTCTATA CCTAGACTCT TTTTGGCAGG AAAGCGACCA	3060
	TCCAGAGGGC CGCTTGCTAG CGCTCTTAAT CCTAGAAATG GGGCTCTTGC CCAGTGAAT	3120
45	TTTAGCCATC AAGGTTGCGG ACATCAATCT GGATTTTCAG GTGCTCCGAA TCAGCAAGGC	3180
	TTCCCAACAG AGGATTGTCA CCATTCCAC GGCCTTGCTT TCAGAATTGG AACCCTTGAT	3240
	GGGGCAGACC TATCTTTTTG AAAGAGGAGA GAAACCTAT TCTCGTCAGT GGGCCTTTTG	3300
50	TCAGTTAGAA TCTTTTGTGA GGAGAAGGTT TCCATCCTTA TCAGCTCAAG TCTTACGTGA	3360
	CAGTTTATTC TAAGCAATAA GAAACAGGTC GATTGTGACG AATTGCAAAA AATTAGGATT	3420
55	AAAAACAGTC CTGACCTTAG AAAATATAGA TAATGGATAT TAAATTAAAA AGATTTTTGA	3480
	AGGACCCCTGG ACTTGCTCTT TGCATCTGGT TTCTAAGTAC CAAGATGGAT ATCTACGATG	3540
	TGCCCATTAC GGAAGTCATC GAACAGTATC TAGCCTATGT TTCAACCCGT CAGGCCATGC	3600
60	GTCTGGAAGT GACGGGTGAG TACATGGTCA TGGCTAGTCA GCTCATGCTG ATTAAGAGTC	3660

WO 98/26072

PCT/US97/22578

-116-

	GTAAACTCCT	TCCGAAGGTA	GCAGAAGTGA	CAGACTTGGG	GGATGACCTG	GAGCAGGACC	3720
	TCCCTCTCTCA	AATCGAAGAA	TATCGCAAGT	TCAAGCTCTT	GGGTGAGCAC	TTGGAAGCCA	3780
5	AGCACCAAGA	ACGGGCCAG	TATTATTCCA	AAGCGCCGAC	AGAGTTGATT	TACGAAGATG	3840
	CGGAGCTTGT	GCATGACAAG	ACGACCATTG	ACCTCTTTTT	GGCTTTTCA	AATATCCTAG	3900
10	CCAGAAAAA	AGAGGAGTTT	GCACAAAAAT	ACACGACGAT	CTTGCGGGAT	GAGTATAAGA	3960
	TTGAGGACAT	GATGATTATT	GTGAAAGAAT	CCTTGATTGG	ACGAGATCAA	TTGCGCTTGC	4020
	AGGATTTGTT	CAAGGAAGCC	CAGAATGTCC	AAGAGGTCAT	CACCTCTTTT	TTGGCAACCC	4080
15	TAGAGTTAAT	CAAAACCCAG	GAGCTGATCC	TGCTGCAAGA	GGAGAGTTTC	GGAGATATCT	4140
	ATCTCATGGA	AAAGAAGGAA	GAAAGTCAAG	TGCTCAAAG	CTAGACTTGA	TAGAGAGGAA	4200
20	AGATGAGTAC	TTTAGCAAAA	ATAGAAGCGC	TCTTGTTTGT	AGCGGGTGAA	GATGGGATTC	4260
	GGGTCCGCCA	GTTAGCTGAA	CTCCTCTCTC	TGCCACCGAC	AGGCATCCAG	CAGAGTTTAG	4320
	GAAATTAGC	CCAGAAGTAT	GAAAAGGACC	CAGATTCCAG	TTTGGCTTTG	ATTGAGACAA	4380
25	GTGGTGCTTA	TAGATTGGTG	ACCAAGCCCT	AATTTGCAGA	GATTTTGAAG	GAATACTCTA	4440
	AGGCGCCTAT	CAACCAAGAG	TTGTCTCGGG	CTGCCCTTGA	GACCTTGTC	ATTATTGCCT	4500
30	ACCAACAGCC	GATTACGGGG	ATAGAAATTG	ATGCCATCCG	TGGGGTTAAC	TCCAGTGGAG	4560
	CCTTGGCAAA	GTTGCAAGCT	TTTGACCTGA	TAAAGGAAGA	CGGGAAAAAG	GAAATATTGG	4620
	GGCGCCCCAA	CCTCTATGTG	ACTACGGATT	ATTCTCTAGA	TTACATGGGG	ATAAACCAAT	4680
35	TAGAAGAATT	ACCAAGTGATT	GATGAGCTTG	AGATTCAAGC	CCAAGAAAGC	CAATTATTTG	4740
	GTGAAAGGAT	AGAAGAAGAT	GAGAATCAAT	AAGTATATTG	CCCACGCAGG	TGTGGCCAGT	4800
40	AGGAGAAAAG	CAGAAGAGCT	GATCAAGCAA	GGTTTGGTGA	CAGTTAACGG	ACAAGTGGTG	4860
	CGTGAACCTAG	CAACCACTAT	CAAGTCAGGC	GACAAGGTCT	AAGTTGAAGG	TCAACCTATC	4920
	TACAACGAAG	AAAAGGTCTA	TTATCTGCTT	AACAAACCAC	GGGTGTCTAT	TTCCAGTGTG	4980
45	ACAGATGACA	AGGGTCGCAA	GACGGTTGTC	GACCTCTTGC	CCAATGTCAA	AGAGCGCAT	5040
	TACCTGTGTG	GTCGTTTGGG	CTGGGATACA	TCAGGAGTCT	TGATTTTGAC	CAATGATGGG	5100
50	GACTTTACAG	ACGAGATGAT	TCACCTCTGT	AATGAGATTG	ACAAGGTTTA	TGTCGCGCGT	5160
	GTTAAAGGTG	TGGCCAAATA	GGACAATCTT	CGCCCCCTGA	CCCGTGTCT	TGAGATTGAT	5220
	GGTAGAAAAA	CCAAGCCAGC	TGTTTATGAA	ATTCTCAAG	TGGACCCAGT	CAAAAATCGC	5280
55	TCTGTGGTGC	AGTTGACCAT	CCATGAAGGG	CGTAACCATC	AGGTTAAAAA	GATGTTTGAA	5340
	GCTGTGGTGC	TCCAAGTAGA	TAAGTTGTCT	CGGACTCGTT	TGGACACCT	AGACTTGACA	5400
60	GACTCCGTCC	AGGGAATCC	CGTCTGCTTA	ATAAAAAAGA	AATCAGCCAA	CTACACACCA	5460
	TGGCTGTAAC	TAAGAAATAA	TGAACCGAAT	TTTAATAGCG	CTTGTGCGCT	TTTACCAACG	5520

-117-

5 TTTTATCTCA CCAGCTCTTC CACCCTCTTG TCGCTTTGAG CTGACTTGT CCAACTACAT 5580  
 GATTGAGGCT ATTGAAAAAC ATGGTTTAA GGGGGTATTG ATGGGCTGG CTCGGATTT 5640  
 ACCTGTGCAT CCCTGGTCGA AAACAGSTAA GGACCCCGTT CCAGACCACT TTTCCCTTAA 5700  
 ACGAAATCAA GAAGGGGAAT GAGGTGGGGT AAATAGATT CAAATGATA AAAACGCATC 5760  
 10 CTATCAGGTT TGAGTGAAC TGAATGAGAT CGTTTATGAA TGTCAAAAT TTATAC 5816

## (2) INFORMATION FOR SEQ ID NO:56:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 725 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

25 (iv) ANTI-SENSE: NO

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:56:

30 TTGAAATAA TTATGAACCG CAATATATTA ATATCCGAGG AAAAGGCCCT CTTATCAATG 60  
 ACTTGAAAAA AGAAGCTAAA AAAGCTAATA AAGTTTTTCT CGCGAGTGAC CCGGACCGTG 120  
 35 AAGGAGAAGC GATTTCCTGG CATTGGCCCC ATATTCTCAA CTGGATGAA AATGATGCCA 180  
 ACCGTGTGGT CTTCATGAA ATCACCAGG ATGCAGTCAA AAATGCTTTT AAAGAACCCTC 240  
 40 GTAAGATCGA TATGACTTG GTCGATGCCC AACAGCTCG TCGGATCTTG GATCGCTTGG 300  
 TAGGGTATTC GATTTCGCCT ATTTTGTGGA AGAAGGTCAA GAAGGGCTTG TCAGCAGGTC 360  
 GCGTTCAGTC CATTGCCCTT AAATCATCA TTGACCGTGA AAATGAAATC AATGCCTTCC 420  
 45 AGCCAGAAGA ATACTGGACA GTTGATGCTG TCTTTAAAAA GGGAAACCAA CAATTCATG 480  
 CTTCCTTCTA TGGAGTAGAT GGTAAAAGA TGAACCTGAC CAGCAATAAC GAAGTCAAGG 540  
 50 AAGCTTGTGTC TCGCTGAGC AGTAAAGACT TTTCAGTAGA TCAGGTGGAT AAGAAAGAGC 600  
 GTAAGGCAAA TGCTCCTTTA CCTATACCA CTTGATCTAT GCAGATGGGA TGCTGCCAAT 660  
 AAAATCAATT TCCGTACTCG AAAAACCATG ATGGTTGCCC AACAGCTCT ATGAAGGAAT 720  
 55 TATAT 725

## (2) INFORMATION FOR SEQ ID NO:57:

60 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1935 base pairs  
 (B) TYPE: nucleic acid

-118-

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

15	AACCCATCTG AAGAACTTT CCGTGCTGCT AGCTCAGCTA TCGATAAAGC AGAAACTAAA	60
	GGTTTGATTC ACAAAAACAA AGCAAGCCGC GATAAAGCTC GTCTTTCAGC TAAACTTGCT	120
20	AAATAAGAAA CAGTCCATAG AGGCTGTTTT TTTGTCTCCA AATAGGAAAA GGTAGAAAAAT	180
	GAAATCACA ATTATCGGAT ATTCIGGTTT TGTTAAGTCA ACTCTAGCAG AAAAGTTATC	240
	TAACTACTAC TCCATTCCAA AACTGCACAT GGACACACTC CAATTTCAAC CTGGTTGGCA	300
25	AGACAGTGAC TGGCAATGGA TGTTAACCGA GATAAAAAAC TTTCTCACC AGCATAAAGC	360
	TTGGGTCATC GATGGTAATT ATTCTTGGTG CTACTACCAA GAACGAATGC AAGAAGCTGA	420
	CCAAATCATC TTTCTCAATT TTTTGCCATT GACCTGTCTC TTAGAGCCT TTAAGCGTTA	480
30	TCTTAAATAC CGTGGAAAAG TCAGAGAAAAG TATGGCCGGC GATTGCCCTG AACGCITTGA	540
	GTGGGAGTTT ATCAGATGGA TTCTTTGGGA TGGGCGTAGC AAAACTCAAA AAGAAAATTA	600
35	CCAAAAACTT TGCCAAGAAT ATTACATAA AGTCACATC CTTCGAAATC AGAGAGAGCT	660
	AGATCAATTT CTGGATAAGA AAAGGAAGTC CTACATTTCA TAAAGGGCTT CCTTTTGGC	720
	TATAATTATT CTGCAATCAA GGTTCCTCAA CCAACCTTCA TCATATCAGT GAAGGTATTT	780
40	TGACGTTCTT CTGCAGTTGT GTCTTCGTCT GGATTGACCA AGCTATCAGA GATGGTCATG	840
	ATAGCTAGCG CATCAACATG GTATTGGGCA GCAAGATAGT AAAGAGCTGC TGCTTCCATT	900
45	TCCACAGCCT TGACTCCCCA TTTACCAAGC TCGATATTCT TTTCAAAGTA ATTTGAGTAA	960
	AAGACATCAG ATGACAAAAC GTTCCCAACG TGAGTAGTCA TACCAAGTTC TTTGGCGATA	1020
	TGGTAGGCTT TATCAAGCAA ATCAAGGCTA GCAATTTGTG GAAATCGTA CTGIGGCCAG	1080
50	TCATTACGAA CGATGTTTGA GTTGGTTGCA GCCGCTGCG CCAAACTAA TTCACGAACA	1140
	TGAACCTCTT CATTCAAAGA ACCTGCAGTT CCCACACGAA TCAATTTCTT CACACCGTAG	1200
55	TCTACGATTA ACTCACGCGC ATAAATCGAA ATAGATGGCA TTCCCATCCC AGTCCCCTAG	1260
	ACAGATACAC GGTGACCTT GTAAGTACCA GTGTAACCAA ACATGTTACG CACTTCGTTA	1320
	AAACAAACAG CATCACCAAG GAAATTCTCC GCAATAAACT TAGCACGAAG AGGATCCCCA	1380
60	GGAGAAGAAA TTTTATCAGC AATTTACCCC TGCTGAGCAG CAATATGGAT AGACATAAAT	1440

PCT/US97/22578

	TATGATACAA AGAGCGAGAA GAAACGACT GAAATATAGG AACCTGACGA GAAATCCTGA	1500
5	TTTTTCAGTC AGATTATCTA TTTTCCGAGT TTTCCGCTCG TGTTCAAATC AAAACACAGG	1560
	CTCTACCTTT CTTTATTTTA TATTTTATAT TGAGAAAGAT ACCAAACCCA TCAAAAAGCG	1620
	AAGGGAAAT AGGAGTTGGG CGCAGTGAGC GATGCTCGCT AGACCAACTA TCTTTTCCCC	1680
10	ACTGCTTTTA GGGTGGGGTC AATTCCTTTC TTTCCTAATT TTGATTTAGA GGAGAGTCGC	1740
	CCGTATTACG TTCACGCAAT ACAGTTTACC CATCCTTTGG TTTTATATTT TAGAAAAGTT	1800
15	TTCTACTCGT GTTCAAATTA GAACACGCGC TCTACCTTTC TGTTTATACT CTCGAAAAT	1860
	CTCTTCAAC ACAGTCAACG TCGACTTGGG TTATATATGT GACTGACTTC GTCATCTTTA	1920
	TTTACAACCT CAAG	1935
20	(2) INFORMATION FOR SEQ ID NO:58:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 2221 base pairs	
25	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA (genomic)	
30	(iii) HYPOTHETICAL: NO	
	(iv) ANTI-SENSE: NO	
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:	
	TATTATTTTT CCCATCTTAA CTGGAACCTA TGTCGCGCGT GTCTTGGACC GAACTGACTA	60
40	TGGTTACTTC AACTCAGTCG ACACATTTTT GTCATTTTTC TTGCCCTTTG CAACCTATGG	120
	TGCTATAAC TACGGTTTAA GGGCTATCAG TAATGTCAAG GATAACAAAA AAGATCTTAA	180
45	CAGAACCTTT TCTAGTCTTT TTTATTTGTG CATGCGTTGT ACGATTTTGA CCACTGCTGT	240
	CTATATCTTA GCTATCTCTC TCTTCTTTAC TGATAATCCA ATCGTCAAAA AGGCTCTACT	300
50	TCTTATGGGG ATTCAACTCA TTGCCCGAGT TTTTTCATTC GAATGGGTCA ATGAAGCTCT	360
	GGAAAATTAC AGTTTTCTCT TTTACAAAAA TGCCCTTCAT CGTATCCTGA TGCTGGTCTC	420
	TATTTTCTTA TTGTGTTAAA ATGAACACGA TATTGTTGTC TATACACTTG TGATGAGTTT	480
55	ATCGACGCTG ATTAACCTAC TGATTAGTTA TTTTGGGATT AAAAGAGACA TCAAACTTGT	540
	TAAAAATGCC CTAAGTGATC TTAACCAACT CTTTCCTCCT CTGACAGCCA TGTTAGTCTT	600
	TGCCAATGCG AATATGCTCT TCACCTTTTT AGATCGCCTC TTCTTCGTTA AAACAGGGAT	660
60	TGTATGCAAC GTTAGTTACT ATACATTAGC TCAGCGAATT GTGACCGTTA TAGCTGGGGT	720

WO 98/26072

PCT/US97/22578

-120-

```

TGTAAACAGGT GCAATTGGAG TGAGTGTGCC TCGTCTCAGT TACTATCTGG GGAAAGGAGA      780
5  CAAAAGAGGCC TATGTTTCTC TGGTAAATAG AGGTAGTCGA ATCTTTAACT TCTTTATCAT      840
TCCACTGAGT TTTGGACTCA TGGTTTTAGG ACCAAATGCC ATCTACTTT ACGGTAGTGA      900
AAAAATATATC GGAGGCGGCA TCTTGACCTC TCTCTCGCT TTCGTACGA TTATCCTGGC      960
10 CTTAGATACC ATTCTTGGTT CCCAAATTCT CTTTACCAAT GGCTATGAAA AACGTATCAC     1020
AGTCTATACA GTCTTGTGCT GGCTACTCAA TTTGGGCTTG AATAGTCTCC TTTTTCCTAA     1080
15 CCATATCGTG GCTCCTGAAT ACTACTTACT GACAACTATG CTATCAGAGA CTTCCTACT      1140
TGTTTTCTTAT ATCATTTTCA TCCATAGAAA ACAACTCATC CACTTGGGAC ATATCTTTAG     1200
CTATACTGTT CGATACTCTC TCTTTTCACT TTCCTTTGTA GCAATTATTT TCCTGATTAA     1260
20 TTTCGTGTAT CCTGTAGATA TGGTCATTAA TTTGCCATTT TTGATTAAAT CTGGTTTGAT     1320
TGCTCTGCTA TCAGATCTCT CTTATATTAG TCTACTTGTC TTCACAAAAG ATAGCATTTT     1380
CTATGAATTT TTAACCATG TCCTAGCCTT AAAAAATAAA TTTAAAAAT CATAGGAGTT     1440
25 TAAATGAAA CAACTAACCG TTGAAGATGC CAAACAAATT GAATTAGAAA TTTTGGATTA     1500
TATTGATACT CTCTGTAAAA AGCACAATAT CAACTATATT ATTAACCTAG GTACTCTGAT     1560
30 TGGGGCGGTT CGACATGAGG GCTTTATCCC TTGGGACGAC GATATTGATC TGTCCATGCC     1620
TAGAGAAGAC TACCAACGAT TTATTAACAT TTTTCAAAA GAAAAAGCA AGTATAAGCT     1680
35 CCTATCCTTA GAAACTGATA AGAACTACTT TAACAACCTT ATCAAGATA CCGACAGTAC     1740
GACTAAAATT ATTGATACTC GAAATACAAA AACCTATGAG TCTGGTATCT TTATCGACCT     1800
GCAGGCATGC AAGCTTGGCG TAATCATGGT CATAGCTGTT TCCTGTGTGA AATTGTTATC     1860
40 CGCTCACAAT TCCACACAAC ATACGAGCGG GAAGCATAAA GTGTAAAGCC TGGGGTGCCT     1920
AATGAGTGAG CTAACTCACA TTAATTGCGT TCGGCTCACT GCCGGCTTC CAGTCGGGAA     1980
45 ACCTGTCTGT CCAGCTGCAT TAATGAATCG GCCAACGCGC GGGAGAGGC GGTTCGCTA     2040
TTGGGCGCCA GGGTGGTTTT TCTTTTTTCA CAGTGAGACG GGCAACAGCT GATTGCCCTT     2100
CACCGCCTGG CCCTGAGAGA GTTGACAGCA GCGGTCCACG CTGGTTTGGC CCAGCAGGCG     2160
50 AAAATCCTGT TTGATGGTGG TTCCGCAAAAT CGGCAAAAT CCTATAAAA TCAAAAGGAA     2220
T

```

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1509 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

-121-

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

5 (iv) ANTI-SENSE: NO

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

TGAATTTTGA	ACAGTACACA	GAATACTAAA	ATATTTCTAG	AAATTAATTT	GAATTTTCTA	60
ATTGGATTTG	TCGCATCTTA	TTTCAATCTA	CTATAGAAAA	AAGTCTTTAA	AATTATAAAA	120
CGCATCATAT	CAAGGTTTTT	CAAAAACCTT	GATATGATGC	CTTTTATTGT	GGGAATATGT	180
ATTTGATTTT	CTACTAAAAA	TATGTTTTTG	AATAACCTCT	ATCTTAGTAG	TTTGTATAAT	240
CCCCCTCAAT	CAGTTTTTGC	GATAAGCTTT	AATGCTATGA	CTATACCACT	CTTGCAATTC	300
TTTTGATGTT	GGTGTCAAT	AATCGCCATA	CATCTGGGTT	AAAAATTGGT	CATATTTTTT	360
GGGAACAGGC	AACATACGGC	CCTCAAACCT	AGTTAAAAAT	AGTCTTTTAA	AGGTATCAAC	420
TGGGAAGATT	TCTTTCATCC	CTTCCTTACC	GATCCCAACT	CCTCCTTCAT	ATTGAGGAGT	480
GTTGGTTACA	GCATTTTGTG	CTAGTTGATC	AATTTTCTTG	TAAAAGTAGC	GAGGATTGAC	540
AAATCGGAGA	GCGTACCAGC	TACATAATCT	AAGAAAAATCT	TTTAGTTTGC	TATCACCGTG	600
AACGTCTCGT	GATTTTTTGA	TATAAGCTAG	TTGACGAAGA	GCCACATACT	TATAGCTCTT	660
GTGCACAATG	CTCAAGTCTG	TAAATCGATC	AATTGGGAAG	ACATCGATGA	AAAGGCTGGT	720
ATCATGACGC	TTGTACTTAA	CATGGTCTTC	TATAACAGTA	GAAGTGTCCT	AAATCGATGC	780
GAATTTATGG	AAGTACCAAG	AAGATGTATC	GTAGGAAGA	ACCTTGTAGC	GAGGGTGATT	840
TTCTTCTTCA	ATAATCTTCA	GTAACGCTC	ATAATCTCTA	CGATAAAGGG	AAATATCAAT	900
ATCATCATCC	CAAGGAATCA	TACCTTTGTG	GCGGATGGCT	CCAAGCATGG	TTCCATAACT	960
GAGAAAAATA	GGAAATATCAT	GTTTCTTACA	AGTCTCATCA	ATATAGTCCA	GCAGGGCTAG	1020
TTGAATTTCT	TTAATTTCTT	TTTTTCTTAA	ATATTGCATC	CTAATCCTCC	AATTTATAAG	1080
CGTGAAATTC	ATGACTGTAG	AAGCGTTTTT	CTTCTGGTGG	TAGGGTCATA	TAATCTCCAT	1140
AAAATTGTGT	CAAATAGTA	TCAAATTTTT	CAGGTGCAGG	AAGGCTTAAA	TTCTCAAAGG	1200
GTAAATCGAT	TGTTTTTACA	AAGGTACCAC	TTGGGAAGAC	TTCTTTTCCC	TTAAATTTTG	1260
AAGGGATAAA	AGCCATATAT	TGCCCATTTT	CACGACTATA	TTTTTGAAAT	TCTTCTCGTA	1320
TTTATTTTGC	AAAATAACGA	GGAGAAACCG	GTGGAAGGAG	TAACCAAGAG	GCTGTTCTGA	1380
TCCAATCTTT	TAAAAGGCTA	TCCTTATAGA	CAATATTTTT	ATGTTTACTG	AAAGACAGCA	1440
GTTTGAAGCT	TTCCAGTTTA	TAAACAAGTAT	CAATGACCTT	AGGATCATCA	AAGCGATCTA	1500

WO 98/26072

PCT/US97/22578

-122-

TAGGGAAAA

1509

(2) INFORMATION FOR SEQ ID NO:60:

- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 671 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

15 (iv) ANTI-SENSE: NO

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

ACAAGGGATT TATCTCTGG GACGACGACC TAGACTTTTT TATGCCTCGT AAAGATTATG 60

25 AGAAATTAGC AGAATTATGG CCTCGTTATG CAGATGAACG TTATTTCTTG TCAAAGAGTC 120

ACAAGGATTT TGTGTATCGT AATCTTTTTA TTACCATTGG TGACAAGAAA ACCACCTGTA 180

TCARGCCTTA TCAGCAGGAT TTGATTTCG CACATGGTCT GGCCTTGGAT GTTTTGCTTT 240

30 TGGATTATTA TCCGAAAAAT CCAGCTGAGC GGA AAAACA GGTTCTGTGG GCCTTGATTT 300

ATTCACCTCT TTGTGCGCAA ACTATTCCAG AAAAGCATGG TGATCTCATG AAATGGGGAA 360

35 GTCCGATTTT ACTGGGTTTG ACTCCAAAA CTCTCCGTTA TCGCATCTGG AAAAAGCTG 420

AGAAAGAAAT GACTAAGTAT GATTTGCGTG ATTGTGATGG CATTACAGAA TTATGCTCAG 480

GTCCCTGGCTA CATGAGAAAC AAGTACCCAA TCACATCTTT TGAAGACAAT CTTTCTTTCG 540

40 CATTTGAAGG AACAGAGATG CCTATTCCAA TCGGCTATGA TGTCTATCTC AGAAGCTGCTT 600

TTGGGGATTA TATGACGCTT CCACGAGCAG ACAAGCAGGT ACCGCATCAT GATACTGTCA 660

45 CTGCTGATAT G 671

(2) INFORMATION FOR SEQ ID NO:61:

- 50 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2766 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

55 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

60



WO 98/26072

PCT/US97/22578

-123-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

	ATCTTATACA AGTCGTAAGC CGCTTCCTTA AAACGAGCTT CTAGTAATTC TTCCAATAAG	60
5	ATAGTAACCT TCACACCAT TGGTGTTCCT AGTGAATAAA GCTGAAAAGC TTGTTCTCCT	120
	TTTGGCAAGT TTGTTCGAA ACGGGCACCT GCTGTGGTC TGTTTAGCCC CGTAAAAGCT	180
10	CCTTGATTAC TAGCTTCATC CTGCCATAGC GTCGGTAATT GATATGCTGA CATCCGAGAC	240
	CTCCCTTAAA TCGCATCTTT GTCAAAACCG AGTTTGCCTT GAATAAATT AACGATTTTCG	300
	ACGATGATA TCATTGAGAA GCTTCCAGCC ATAACAATTC CCCATTGTGA CAAGCTTAGT	360
15	TTGGTTACGT GGAAGATTCC TTCAAGCGGT TCTACAACGA TTGTTGCCAT GAGAAGGATA	420
	AAGGATACCA AGATGGACCA GTTAAAGGTC TTAGACTTGA ATGGGCCAAC TGTCAGATG	480
20	GATTGGTAGA CAGACTTGAC ATTGTAGGCA TGAAGAGCT GAATCAAACC AAGGGTTGCA	540
	AAGGCCATCG TTAGGGCATC TGCATGAATA GCATGATTGT CACCCACATG AACGGGTAA	600
	GCAATCGCAA GGGCATAAAC ACTCATAACA AGAGCTGCTT GGAGTACACC TTGATAAATG	660
25	ATAGAACTCA AAACACCACC TGAGAAGAAG CTTCCTTTGC GTCCACGTGG TTTATGATTC	720
	ATGACACCAG GTTCCGCGG TTCAACACCA AGAGCGATAG CTGGGAAGGT ATCCGTTACC	780
30	AAGTTGATCC ACAAAGATG AACCGGCTGC AAGACATCCC AACCAAAACA GGTTGATAGG	840
	AAGATGGTTA ATACTTCAGC AGTATTAGCA GAAAGTAGGT ACTGAATAGT CTTTGAATG	900
35	TTTGAGAAGA CCTTACGTCC TTCTTCCACT GCGACGATA TAGTCGCAA GTTATCATCT	960
	GCAAGATCA TATCAGAAGC CCCCTTAGAA ACCTCTGTAC CAGTGATTCC CATACCGATA	1020
	CCAATATCGG CTGTTTTCAG AGCTGGCGCG TCATTGACAC CGTCACCTGT CATGGCAACG	1080
40	ACTTTACCTT GTTTTTGCCA AGCCTTGACG ATACGAACCT TGTGTTCTGG AGACACACGG	1140
	GCATAACAG AGTATTGACC AACAACTTTT TCAAAATCTT CATCTGACAG TTCATTGAGT	1200
	TCAGCACCAG TTAAACAGTG ACCTTCTGTA TCGTTTGCCT CAATGATTCC CAAACGTTTG	1260
45	GCAATGGCTT CCGCTGTGTC TTGGTGGTCA CCTGTAATCA TAATTGGACG GATTCCCGCT	1320
	TCCTTAGCCA CACGAACAGC CTCAGCGGCT TCAGGAGCTT CAGGCTCAAT CATCCCAATC	1380
50	AAACGATGAA AAATTAAATC ATTTTCAAGC TCTTCAGAG TGAGATTTC TGGAATACTA	1440
	TGATAATCT TATAAGCACC TGCAAGGACA CGCAAGGCTT GATGAGCCAT TTCAGAATTG	1500
55	TTTGTATGAA TGAGATTGT AACCTTCTCA TCAATCGGAG CAATATCCCC AGCCTTATCA	1560
	CGAAGAAGAC AACGTTTAA GAGTTGGTCT GCGCAGCCCT TGACTGCTAC AAGGAAACGA	1620
	CTATCTGGCA ATGGGTGAAC TGTTGACATG AGCTTACGGT CAGAGTCAAA TGGCAATCA	1680
60	GCTACACGAG GATATTTCTC TAAGAAACCT TTGACATCAT AGCCCTTGTC CAAGGCATAT	1740

WO 98/26072

PCT/US97/22578

-124-

```

TGGATAAAGG CTGTTTCGGT TGGGTCACCA ATCAAGTTAC GTTCCACATC GATTTTCGTA 1800
TCATTGGCCA AGACAACTGA ACGAAGTAGT GGCATTTCAA GACCTAGTTC AATATCATCA 1860
5 GCTGAGTCAT GTAGAACCGC ATCGTAGAAG ACTTTTTCGA CTGTCACTTT GTTCATAGTC 1920
AGCGTACCAG TCTTATCAGA AGCGATGATT TCAGTTGAAC CAAGTGTTTC AACTGCTGGC 1980
10 AACTTACGAA CGATGGAATG TCGTTTGGCC AAAACTTGAG TACCAAGAGA AAGAACGATG 2040
GTAACGATAG CAGGAAGTCC TTCTGGAATG GCTGCAACAG CAAGGGCAAC AGAAGTCAAC 2100
AACTCACCAA GTGGATTTTT CCCTTGAATG AAGACACCCA CTACAAAAGT AACAGGGCA 2160
15 ATGACCAAGA TAGCATAGGT CAAGACCTTA GAAAGTTGT TCAAGTTTTG TTTAGTGGT 2220
GTATCAGTCT CATCCGATC TTGAAGCATA CCAGCAATAT GACCAACTTC AGTATACATA 2280
20 CCTGTATTGA CAACAACACC CATCCACGA CCATAGGTTA CGTTTGAGTT TTGGAAGGCC 2340
ATGTTGACAC GGTCAACAA GTCCAGCATCT GTCCGAAGAT CGACTGACAA GTCTTTTTCG 2400
ACTGTTACAG ATTCACTGT CAAGGCTGCT TCTTCAATTT TAAGAGAGTT GGCTTCTATC 2460
25 AAACGTAGGT CCGCTGGTAC CACGTCACCT GCTTCAAGGG CAACGATATC GCCTGGTACC 2520
AATCTTTTGA AGTCAATCTC TGCCATGTGT CCATCACGAA GAACGCCGGC AACTGGACTA 2580
GACATGGATT TGAGGGCTTC AATAGCTTCT TCAGCTTTTC CTTCCTGGTA AACACCAAAG 2640
30 GCAGCGTTGA TGATAACCA AGCTAGGATG ATAATGGCAT CTGCGATATC TTCCCACCA 2700
GAAGTCACGA CTGACAAGAT TCTGCCGCAA CTAGGATGAT AATCATCAA TCCTTAAATT 2760
35 GCTCGA 2766

```

(2) INFORMATION FOR SEQ ID NO:62:

```

40 (i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 1577 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: DNA (genomic)

(ii) HYPOTHETICAL: NO

50 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

55 TGGATTTATC CTCTTTTTCG TTCTTTTGGG AGCAGTTTTT GAGGAAAAAA TGAGAAAAAA 60
TAGTCCCAAA GCTGTGGAGA AATTACTGGA CTGCAAGCT AAAACCCGAG AAGTCTTGAG 120
60 TGATGATAGT TATGTCCAAG TTCCTTTGGA ACAAGTCAAG GTAGGCGACC TGATTCGAGT 180

```

PCT/US97/22578

	GCGTCCCGGT	GAAAAGATTG	CTGTTGATGG	TGTCGTAGTA	GAAGTGTCCT	CTAGATTTGA	240
	CGAATCCATG	GTGACAGGTG	AGAGTCTGCC	TGTGGACAAG	ACAGTTGGAG	ATACTGTCACT	300
5	TGGCTCAACC	ATCAATCATATA	GTGGAAACGCT	TGCTCTTTAGA	GCAGAAAAAG	TTGGCTCAGA	360
	GACTGTTTGG	GCTCAGATTG	TGGATTTTGT	GAGAGAAAGCT	CAGACCAAGTC	GTGCGCCGAT	420
10	TCAGGACITG	ACGGATAAGA	TTTCAGGGAT	TTTTGTCCCA	GTAGTTGTCA	TTTTAGGAAAT	480
	CATGACCTTT	TGGSTTTGGT	TCGTCCTTGT	CAGGGATAGT	GTGCTCGTGC	TTGGAGCTAG	540
	CTTTGTGTCC	TCTCTTCTCT	ACGGAGTGGC	GGTTTGATTA	TCGCCCTGTC	TTGTGCCTTG	600
15	GGACTTGCAA	CACCGACAGC	CCTTATGGTG	GGGACAGGAC	GTAGTGCCAA	GATGGGGGTT	660
	CTCCTCAAAA	ATGGAACGTG	CTTACAGGAA	ATCCAGAAAG	TTCAAACCTG	TGCTCTTTGAT	720
20	AAGACCGGGA	CTTTGACCGA	AGGGAAACCT	GTGGTAACAG	ATATCATCGG	CAGCAGGAATG	780
	GAAAGTGTGG	GATTGGCAGC	TCTCCTTGAA	GATGCTTCTC	AACACCCACT	GGCTGAGGGCT	840
	ATCGTTAAGC	GAGCGAGTGA	AGCTGGACCT	GAGTTTCAAA	CTGTTGAAAA	TTTTCAGGCC	900
25	TTGCACGGGA	AAGGTGTTTC	AGGGCGAATC	AATGAAAAAC	AAGTTTACTT	TGGAAATGCT	960
	AAAATGCTGG	ATGGCATGGA	TATTTCTAAT	ACTTATCAAG	ATAAACTAGA	AGAACTAGAA	1020
30	AAAGAAGCTA	AGACAGTGTG	GTTTTTAGCT	GTTGACAATG	AAATCAAAGG	CTTGCTTGCT	1080
	TTGCAAGATA	TTCTTAAGGA	AAATGCTAAG	CTAGCCATCA	GTCAGCTAAA	AAAACGTGGT	1140
	CTCCGAACAG	TCATGCTGAC	AGGAGACAAT	GCTGGGTGGG	CGCGTGCTAT	TGCAGATCAA	1200
35	ATCGGAATTG	AAGAGGTGAT	TGCAGGCGTC	TTGCCAAGAG	AAAAAGCCCA	TGAAATCCAT	1260
	AAACTGCAAG	CGGCTGGCAA	AGTAGCCCTT	GTGGGGGACG	GTATCAATGA	CGCTCCTGCC	1320
40	CTTAGTGTAG	CAGATGTGGG	AATTGCTATG	GGTGCTGGAA	CAGATATCGC	CATCGAGTCA	1380
	GCAGATTGGG	TGTTGACAAC	CAATAATCTT	TTAGGAGTGG	TTCGTGCCTT	TGATATGAGT	1440
	AAGAAAACCT	TTATCAGAAAT	TCTACTCAAT	CTTTTCTGGG	CTTTTATCTA	CAATGTTGTG	1500
45	GGAAATTCCGA	TTGACGAGAT	AGTCTTTTCA	GGTGTGTGGT	GGCTCTCAAC	CCAGATTGGC	1560
	AAGGATAGCC	CAATGGC					1577

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1089 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

-126-

5           (x1) SEQUENCE DESCRIPTION: SEQ ID NO:63:  
AAAAATGATAT AATAGAATTT ATGGATAAAA ATAAGATTAT GGGATTAAACC CAAAGAGAAG       60  
TCAAGGAAAAG ACAGGCTGAG GGTTTGGTCA ATGACTTTAC CGCATCAGCC AGTACCAGCA       120  
10 CTTGGCAAAT CGTTAAACGA AATGCTTTTA CCTTTTTTAA CGCTTTGAAC TTIGCCATTG       180  
CTTTGGCCCT TGCTTTTG TGAGGCTTGA GCAATCTGGT CTTCTTTGCT GTTATCTGCT       240  
15 TTAACGCTTT TTCTGGGATT GTGACCGAGC TACGAGCCAA ACACATGGTG GACAAGCTCA       300  
ATCTCATGAC CAAGGAAAAG GTCAAAACCA TCCGTGATGT CAGGAAGTTG CTCCTTAATCC       360  
20 TGAAGAATTA GTGCTAGGAG ATGTATTTCG TTTGTCTGCA GGAGAGCAGA TTCCTAGTGA       420  
TGCCTTGGTT TTGGAAGGCT TTGCGGAAGT CAATGAAGCC ATGTTAACGG GAGAAAGTGA       480  
TTTGGTGCAA AAGGAAGTTG ACGGCTTACT TTTGTCAGGA AGTTTCCTAG CCAGTGGGTC       540  
25 AGTTTATCT CAACTTCACC ATGTCGGTGC AGACAACTAT GCTGCCAAAC TCATGCTTGA       600  
GGCTAAGACC GTTAAACCCA TCACTCCCG TATCATGAAA TCGCTGACCA AGTTGGCTGG       660  
30 TTTTACTGGG AAGATTATCA TTCCCTTTGG TCTGGCTCTC TTGCTGGAAG CCTTGCTTTT       720  
AAAAGGCCTG CCTCTCAAGT CATCCGTTGT AAATCGTCTG ACAGCTCTTT TGGGAATGTT       780  
GCCTAAGGGA ATTGCCCTTT TGACCATTAC TTCGCTCTTG ACTGCAGTGA TTAAGTTGGG       840  
35 CTTGAAAAAG GTCTGGTGC AGGAGATGTA CTCTGTTGAG ACCTTGGCGC GCGTGGATAT       900  
GCTCTGTCTG GACAAGACGG GCACCATCAC CCAAGGAAAG ATGCAGGTGG AGGCTGTTCT       960  
40 TCCACTGACG GAAACTTACG GTGAAGAGGC TATTGCCAGC ATCTTGACTA GCTACATGGC       1020  
CCATAGTGAG GATAAGAATC CAACTGCCCA AGCCATTGCG CAGCGTTGTG GGGAGATGTT       1080  
GCTTATCCT       1089

45           (2) INFORMATION FOR SEQ ID NO:64:  
          (i) SEQUENCE CHARACTERISTICS:  
            (A) LENGTH: 731 base pairs  
            (B) TYPE: nucleic acid  
50           (C) STRANDEDNESS: single  
            (D) TOPOLOGY: linear  
          (ii) MOLECULE TYPE: DNA (genomic)  
55           (iii) HYPOTHETICAL: NO  
          (iv) ANTI-SENSE: NO

60

WO 98/26072

PCT/US97/22578

-127-

(x1) SEQUENCE DESCRIPTION: SEQ ID No:64:

	GCTAGCAATA TCATGTTTAT GCTTGATTG GGAATCATT TAGATCAGTG GTCCTTGAAA	60
5	AAAAC TGCAA CAGATTAGA ACAGAGCTTT CTTGCAAAAG AGAGCGATGT ATTCTTAGTA	120
	CAGGGCGATA CGGTTGTTAG TATCAAGAGT TCCGATGTTT AATAGGAGA TGTCTTGATC	180
10	TTATCTCAAG GAAATGAAAT TCTGTTTGAT GGACAAGTAG TTTGAGTTT AGGTATGGTC	240
	AACGAAAGCT CTTTGACAGG AGAGAGTTTT CCAGTTGAAA AAAGAGAGTC TGATTTGGTT	300
	TGTGCAATA CAGTATTAGA AACTGGAGAG TTACGCATTG GTGTAAACAGA TAATCAGATG	360
15	AACAGCCGTA TTTTACAGCT GATTGAGTTG ATGAAGAAAT CTGAAGAAA CAAGAAAACG	420
	AAACAACGCT ATTTTCATCA GATGGCGGAT AAAGTCGCA AATATAAATT CTTGGGCTCT	480
20	GGGCTGACTT ACCTATTGAC AGGTTCTTTT TCTAAGGCTA TTTCTTCTCT ATTGGTCGAT	540
	TTCTCCTGCG CTTTGAAAA CTCTACTCCT GTAGCTTATT TGACAGTTAT CAAGGTAGGG	600
	TTGAACCGTG AAATGGTGTG TAAGGATGGA GATGTTCTGG AGAAATATCT GGTAGTTGAT	660
25	ACTTCTTGTG TTGATAAGAC AGGACCAATC ACACTAGTT ATCCTATAGT TGAAGAGGTG	720
	TACCTTTGG G	731

(2) INFORMATION FOR SEQ ID No:65:

30	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 2197 base pairs
	(B) TYPE: nucleic acid
35	(C) STRANDEDNESS: single
	(D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: DNA (genomic)
40	(iii) HYPOTHETICAL: NO
	(iv) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID No:65:

	TATATTATTC CATTGTGGT AAATCTGTAC ATGATAGATT AAGTACTCCG ACTGAAACCA	60
50	GTACACTAAT CAAGCTATAG CCAGCTAACA AAAGGAGTAA CCATAGAATA TTAACCTTTA	120
	AAATTTCCCTT CATCGTTTAC ACCTTCTCTT TCACATTCTT ACCAAGGATA CCAGCTGGGC	180
55	GGACAATCAA GATCAACAAC AAGATTCCAT AAACAATGGC ATCACGGAAA TCTGACATCC	240
	CAAGGCTGTG CGCAAAGGTT TCCAATAGAC CAATCACAAA GCCACCAAGA GCCGACCAG	300
	GAAATAATTCC GATACCACCA AGTACTGCGG CAACGAAAGA TTTAAGACCT GGAGTAACCC	360
60	CCATCAAAGG CTCAAGAGAG TTATAATAAA GAGCAATCAG AACACCAGCC GCACCCGCAA	420

WO 98/26072

PCT/US97/22578

-128-

	GAGCAGAAC	CAAAGCGAAG	GTAAAGCTAA	TGCTACGGTT	TATATTGATC	CCCATCAATT	480
	GGCGCGGTC	GCTATCTACT	GATACTGCAC	GCATGGCTTT	CCCCATCTTA	GTCTTTTGGG	540
5	CAATGACTTG	TAACAAAATC	ATCAAAATCA	AGGAAATGCC	CAAAATCAAT	AATGCGACAT	600
	TTGTTAAGCT	AATTGGTCCC	AAATCATATC	GAAGCTTTTG	AATCGCTTGA	GGGAAAGCAC	660
10	GGGTATTGGC	ACCAACCCAGA	TAGACCATTG	CATACTCCAA	TAGGAAAGAA	ACCCCAATAG	720
	CCGTAATCAA	AACGACCAATA	CGAGTAGAGT	GGCGCAAAGG	TCGGTAAGCA	AGAAACTCAA	780
	TCACGACACC	AAGAAATAGCT	GTGCTAGCA	TAGCTACAA	AAGCGCTACA	AAGAAATCCA	840
15	TTTGGAAGA	ATTGATCAAG	AAATAACCGA	TAAAGGCTCC	CATCATATAA	ATATCACCAT	900
	GGCGGAGTT	GATGAGCTTG	ATAATTCCGT	AAACCATGGT	ATATCCTAGG	GCTAACAGCG	960
20	CGTAAACACT	ACCTAGAATC	AAACCATTTA	CGAGTTGTTG	GAGCATAAGA	TTCACTCTTT	1020
	CTATTATATA	TTCCGAGGGT	TTTCCCTCAC	TTTTTGATAG	GTCTTATAC	TCAATGAAAA	1080
	TCAAAGAGCA	AATCAGGAAA	CTAGCCGCGC	GTGCTCAAA	GCAGTCTGTT	GAGGTTGTAG	1140
25	ATAAGACTGA	CGAATCAGT	CACATATATA	ATCCAAGGCG	ACGTTGACGC	AGTTTGAAGA	1200
	GATTTTCGAA	GAGTATTAAA	TATCGAAACA	GGGAGTGAGT	CAAAGGCTCA	TTCCCTATTT	1260
30	CAACATTTTT	CTATTATGGT	TTTCAAACTT	CTGCTGCTTC	AACTTTACCA	TTGTTCAATGG	1320
	TCATCATGTA	AGCAGTTTTG	ACTGTGTTGT	GGTCTGCATC	GAAGCTTGT	TGACCAGTTA	1380
	CACCTTCAAA	ATCTTTTGT	TTAGCAAGGT	TATCTTGAT	TTCACTGAA	TTTTTAGCAC	1440
35	CTTTTGCTGC	GTGTTGCTACA	AGGTGAACCT	AATCATAAGC	CAAGGCTGCA	AATGTTGAAG	1500
	GCTCTTCATT	GTAATCTAGCA	CGGTAAAGCT	CAAGGAAGGC	TTTAGCTTTA	GCTGAACTT	1560
40	GTACAGTAGT	TGAGAAAGCT	GAGATAAAGT	AGATGTTTGA	TGCTTTTCCA	GCAGTTGCTT	1620
	GTGTACAAA	CTCCTCAGC	TTGAATCCAT	CACCACCAAC	GATTGGTTTG	TCAATTCCCA	1680
	TACCACGGCG	TTGGTTTACA	ATCTTACCAG	CCTCATTATA	GTAAACAGGA	ACAACGATAG	1740
45	CATCAAAGTC	TTTCCCTTTC	ATTTTGTATA	GGGCTGCTTG	GAAGCTGTG	TCACCTGCTA	1800
	CGAAAGTTTC	ATCTGCAACG	ATTTCAACCT	TGTATGACTC	CGGAAAGAT	TTGGCAATCC	1860
50	CTTTAGCATA	GTCACTGGCA	TTGTCAGTGT	AAAGAACAC	TTCTTAGCA	TTTAATTTTT	1920
	CAGAAACATA	GTTTGAGATA	ATTTTCCCT	GGAAGCTATC	TTGGAAAGTT	CCAAATAAGA	1980
	GGTAATCTTG	ACCTTTAGTC	AATCCATCTT	GAGTCGCACT	TGGTGAGATC	AATGGAACAC	2040
55	CTGCTTTTGT	AGCGTTCCGT	ACCGCAGCTG	CAGTCGCACC	AGATGTCGCA	GGTCTACGA	2100
	CTGCTGATAC	TTTAGATTGG	GTTACAAAGT	TAGTTGTAAC	TGAAGCAGCC	TCAGCTGTTT	2160
60	CAGACTTATT	ATCTTTATCG	ACTACTTCGA	TTTGTGTT			2197

(2) INFORMATION FOR SEQ ID NO:66:

PCT/US97/22578

```

5      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 900 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: DNA (genomic)

10     (iii) HYPOTHETICAL: NO

      (iv) ANTI-SENSE: NO

15     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

20     TGTCCCAAGA  CGAGACTTGG  TATGCTCTGG  CCTATGATGG  GGCAGAAGTG  ATTGGCTTTC  60
      TAACGTGTCA  GGAGACTCTC  TTTGAAGCAG  AAGTCTGTCA  AATCGCTGTC  AAAGGAGCTT  120
      ATCAGGGTCA  GGGAAATTGG  TCAGCCTTGT  TTGCTCAATT  GCCGACAGAC  AAGGAAATTT  180
      TCCTCGAAGT  CAGACATGAT  AACTCAACGAG  CGCAAGCATT  TTACAAGAAA  GAAAAGATGT  240
      CAGTTATCGC  TGAGCGGAAG  GCTCTACTAC  ATGACCCAGT  CGAGGACGCC  ATTATCATGA  300
      AGAGAGAAAT  AGATGAAGGA  TAGATATATT  TTAGCATTGG  AGACATCCTG  TGATGAGACC  360
      AGTGTGCGCG  TCTTGA AAAA  CGACGATGAG  CTCTGTGCCA  ATGTCTATTG  TAGTCAAATT  420
      GAGAGTCACA  AACGTTT TGG  TGGCGTAGTG  CCCGAAGTAG  CCACTGCTCA  CCATGTCGAG  480
      GTCAATTACAG  CCGTGTATCG  GGAGGCATTG  GCAGAAGCAG  GGATTACCGA  AGAGGACGTG  540
      ACAGCTGTGG  CGGTTACTCT  CGGACCAGGC  TTGGTCGGAG  CCTTGCTAGT  TGGTTTGTCA  600
      GCTGCCAAGG  CCTTGTGCTG  GGTCTACGGA  CTTGCACCTG  TTCTGTTTAA  TCACATGGCT  660
      GGGCACCTCA  TGGCAGCTCA  GAGTGTGGAG  CCTTTTGGAG  TTTCCCTTGC  TAGCCCTTTT  720
      AGTTCAGTGG  GTGGGGGACA  CAGAGTTGGT  CTATGTTTCT  GAGGCTGGCG  ATTACAAGAA  780
      TTGTTGGGGA  AGACACGAGA  CGATGCAGTT  GGGGAGGCTT  ATGACAAGST  CGGTCGTGTC  840
      ATGCGCTGAC  CTATCCTGCA  GGTCTGAGA  TTGACGAGCT  GGCTCATCAG  GGGCAGGATA  900

50     (2) INFORMATION FOR SEQ ID NO:67:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 1023 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: DNA (genomic)

10     (iii) HYPOTHETICAL: NO

```

WO 98/26072

PCT/US97/22578

-130-

(iv) ANTI-SENSE: NO

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

	CCGGCGATCT TCCGCTAGAA ATAGTCTACC AAGATGAGGA TGTGGCTGTC GTTAAACAAAC	60
10	CTCAGGGAAA TGGTTTGCA CCCGAGTGCT GGTATACCA GTGGAACCTT AGTAATGCC	120
	CTCATGTATC ATATTAAGGA CTGTGCGGGT ATCAATGGGG TTCTGCGTCC AGGGATTGTT	180
15	CACCGTATTG ATAAGGATAC GTCAGGTCTT CTCATGATTG CTAATAACGA TGATGCGCAT	240
	CTAGTACTTG CCCAAGAACT CAAAGATAAA AAGTCTCTCC GCAATATTG GGCATTGTT	300
	CATGGAAATC TGCCTAATGA TCGTGGTGTA ATTGAAGCGC CGATTGGCCG GAGTGA AAAA	360
20	GACCGTAAGA AACAGGCTGT AACTGCTAAA GGAAGCGCTG CAGTGACGCG TTTTCACGTC	420
	TTGGAACGCT TTGGCGATTA TAGCTTAGTA GAGTTGCAAC TGGAGACAGG GCGCACTCAT	480
25	CAATCCGTG TCCACATGGC TTATATCGGC CATCCAGTCG CTGGTGATGA GGTCTATGGT	540
	CCTGCAAGAC TTGAAAGGA CATGGACAAT TTCTTCATGC CAAGACTTTA GGTTTTACTC	600
	ATCCGAGAAC AGGTAAGACC TTGGAATTA AAGCAGATAT CCCAGAGATT TTTAAGGAAA	660
30	CCTTGGAGAG ATTGAGAAAG TAAGAAATGA AAAGAAATTA ACTAGTTTAC CACTTGTAGG	720
	CGCTTTTTTA GGTTTGTCAT GGTATGGGAA TGTTCAGGCT CAAGAAAGTT CCAGGAAATA	780
35	AAATCCACTT TATCAATGTT CAAGAAAGTG GCAGTGATGC GATTATTCTT GAAAGCAATG	840
	GACATTTTGC CATGGTGGAT ACAGGAGAAG ATTATGATTT CCCAGATGGA AGTGATTCTC	900
	GTTATCCATG GAGAGAAGGA ATTGAACCTT CTTATAAGCA TGTTCATAACA GACCGTGTCT	960
40	TTGCTCGTTT GAAGGAATTG AGTGTC AAAA AACTTGATTT TATTTTGGTG ACCCATAACC	1020
	ACA	1023

(2) INFORMATION FOR SEQ ID NO:68:

45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 675 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

50

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

55

(iv) ANTI-SENSE: NO

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:



WO 98/26072

PCT/US97/22578

-131-

	GCTCGGTACC CGGGGATCCT CTAGAGTCGA TAATATCAAC CTGCAGGTTG ATGAACGAGA	60
5	TCGGATTGCT CTTGTTGGGA AAAATGGTGC AGGTAAGTCT ACTCTTTTGA AGATTTTAGT	120
	TGGAGAAGAG GAGCCAACTA CGGAGAAAT CAATAAGAAA AAAGATATT CTCTGTCTTA	180
	CCTAGCCCAA GATAGCCGTT TTGAGTCTGA AATACCATC TACGATGAAA TGCTTCATGT	240
10	CTTTAATGAT TTGCTGCGGA CGGAGAGACA ACTGCGTCAG ATGGAGCTGG AGATGGGTGA	300
	AAAGTCTGGT GAGGATTTGG ATAACTGAT GTGAGATTAT GACCGCTTAT CTGAGAATTT	360
15	TGCGCAAGCA GGTGCTTTTA CCTATGAAGC TGATATTCGA GCGATTTTGA ATGGATTCAA	420
	GTTTGACGAG TCTATGTGGC AGATGAAAAT TGCTGAGCTT TCTGGTGGTC AAAATACTCG	480
	TTTGCACTTT GCCAAAATGC TCCTTGAAA GCGCAATCTC TTGGTCTGG ACGAGCCCAAC	540
20	TAAACACTTG GATATTGAAA CCATCGCCTG GCTAGAGAAAT TACTTGTTAA ACTATAGCGG	600
	TGCGCTCATT ATCGTCAGCC ACGACCGTTA TTCTTGGAC AAGGTTGCGA CAATTACGCT	660
25	AGATTTGACC AGCAT	675

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 582 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

45	TAGAGTCGAT AGCAATAGAT TGAGTAAGTT GTAGCCTTAC CAGCATCGAT AGAAACAAGG	60
	GCACCACGGT GACGTCACC AATTCCCCCT GGAATCAATG GCAAGTATTG GTGCAAGGTA	120
50	TGGTTCATGA TACCGTAACC ACGAGTCATT GACAAGAACT CAGTTGAGTA TCCAATCAA	180
	CCACGCGCTG GAACAAGGAA GACCAACGA GTTTGACCAT TACCAGTTGA AATCATATCC	240
	AACATTTCAC CTTTACGTTT AGAAAGGCTT TGGATAACAG ACCCTTGGTA TTCTTCTGGA	300
55	GTGTCGATTT GTACACGTTT AAATGGTTCA CATTTAACAC CGTCGATTTT TTTTACGATA	360
	ACTTCTGGAC GAGATACCTG AAGTTCATAG CCTTCACGAC GCATTGTTGT GATAAGGATT	420
	GACAAGTGCA ATTCTCCACG TCCTGAAACA GTCCATTTAT CTGCGTGAAT CAGTTGGGTC	480
60	AACACGAAGG AACGTCTGTT TGCAATTCTG CCTGCAAGCG TTCTTCCACC TTACGAGAAG	540

-132-

TTACCCATTT ACCTTCTTTA CCAGCAAAATG GTGAGTTGTT GA

582

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1337 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

TTGGATTGAA	GAACAAAGAT	TTGGACTCTA	TTGACCTTAT	GGTTGGGGG	AAATTTGGAA	60
TTTCAAAGTC	GCCCAACCCC	CTCATTTTCTA	AAGAATTGGA	AGCCGGATGG	GACTCTACCA	120
AACACGTTTA	ACCCAAAGAA	AATTGGCAGG	AAGAAATGGA	AAAAATTGGA	TTTTTAAAAA	180
ATACTTAAGG	AACTTTAAG	CTAGGGAGTG	TACCCTAAGT	TCAATAAGT	TAAAGAAGAC	240
CTTAACTTAA	ACTCCTAAAA	CTTTTCAAT	AATAATCTCC	CTATAAAAAA	AAAGTCGCCC	300
AATCAGGCGG	CTTAATTTTT	TTGAAAAATG	GGCTTGGTGC	CTGAGAATAA	ATAGCTTAGT	360
GATAGAAGAA	AATGGGGAAA	TATGTATATA	TGAAACGATA	GATTTTGGAA	TAGGAATAAG	420
ATCATGTTTG	GATTTTTTAA	GAAAGATAAG	GCTGTGGAAG	TAGAGGTTCC	GACACAGGTT	480
CCTGCTCATA	TCGGCATCAT	CATGGATGGC	AATGGCCGTT	GGCTAAAAA	ACGTATGCAG	540
CCGCGAGTTT	TTGGACACAA	GGCGGGCATG	GAAGCAITGC	AAACCGTGAC	CAAGGCAGCC	600
AACAAACTGG	GCGTCAAGGT	TATTACGGTC	TATGCTTTTT	CTACGGAAAA	CTGGACCCGT	660
CCAGATCAGG	AAGTCAAGTT	TTCATGAACT	TGCCAGTAGA	GTTTTATGAT	AATTATGTCC	720
CGGAATACA	TGCGAATAAT	GTTAAGATTG	AAATGATTGG	GGAGACAGAC	CGCCTGCCTA	780
AGCAAACTTT	TGAAGCTTTA	ACCAAGGCTG	AGGAATTGAC	TAAGAACAAC	ACAGGATTGA	840
TTTCCACAGG	TGTTTTAGAT	GCCAAAATCA	ACCCAGGTGA	CATCACAGAG	GAATTGATTG	900
GTAACATATCT	CTTACTACAG	CATTTCGCTA	AGGACTTACG	AGACCCAGAC	TTGATTATCC	1020
GTACTAGTGG	AGAATTACGT	TTGAGCAATT	TCCTTCCATT	GCAGGGAGCC	TATAGTGAGC	1080
TTTATTTTAC	GGACACCTTA	TGGCCTGATT	TTGACGAAGC	GGCCTTGCAG	GAAGCTATTC	1140
TTGCCTATAA	TCGTCGTCAT	CGCCGATTTG	GAGGAGTTTA	GGAGGAATAA	TGACCCAGGA	1200

WO 98/26072

PCT/US97/22578

-133-

TTTACAGAAA AGAACCTTGT TTGCAGGGAT TGCCCTGGCT ATTTTCTTAC CAATTTTAAT 1260  
 5 GATTGGGGGC TCTTGCTTCA GATAGCAATC GGAATCCTAG CCATGCTAGC CATGCATGAA 1320  
 CTTTGAAGA TAAGAGG 1337

(2) INFORMATION FOR SEQ ID NO:71:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 818 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

20 (iv) ANTI-SENSE: NO

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

TCGGTACCCG GGGATCCTCT AGAGTCGATA GTCGCCAAGC AGAAGAAGGG AACACCATTC 60  
 GTAGAAGACG TGAGTGGGAC GAATGCCAAC ACCGTTTTAC AACCTACGAA CGAGTAGAAG 120  
 30 AAAGAACCCT AGTGGTTGTT AAAAAAGATG GCACACGGGA ACAATTCTCC AGAGATAAAA 180  
 TCTTTAATGG GATTATCCGC TCAGCCGAGA AACGTCCTGT GTCAAGTGAT GAAATCAACA 240  
 35 TGGTGATCCT CTAGAGTCGA ACAGAAACTC CGTGGTCGAA ATGAAAATGA AATTCAAAGT 300  
 GAGGACATTG GTTCACTCGT CATGGAGGAG TTGGCTGAAT TGGACGAGAT TACCTATGTA 360  
 CGTTTTGCTA GTGTCTATCG TAGTTTTAAG GATGTCAGTG AGTTAGAGAG CTTGCTCCAA 420  
 40 CAAATCACCC AGTCCTCTAA AAAGAAAAAG GAAAGATAAA TGAAGCCAAT TGACCGTTTT 480  
 TCTTATCTAA AGAATAATCG GGTGTCGCAA GATACCTCAT CTCTGGTACA GTGCTACCTC 540  
 45 CCGATTATCG GTCAGGAGGC ACTGAGCCTT TATCTTTATA CGATTAGTTT TTGGGATAAT 600  
 GGTAGAAGG AATATCTTTT TTCAAGTATC CTCATCATC TTAACTTTGG AATGGATAGA 660  
 CTGATAAAAT CATIGAAAAA CTTATCTGCT TTTAATCTCT TGACTCTCTA TCAAAAGGGG 720  
 50 GATGTTTATC AGCTAGCCCT CCATGCTCCT CTATCTAGTC AAGACTTCTT GGGGCATCCT 780  
 GTTTATCGCA GACTCTTAGA GAAAAAGATT GGGGACAA 818

55 (2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 746 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 60 (D) TOPOLOGY: linear

WO 98/26072

PCT/US97/22578

-134-

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

TTACCCGGGG	ATCCTCTAGA	GTGATATGC	TCTCTGAGGG	TCAATTCCTC	ATACAGACTA	60
GGCGTCTCAG	GAATGTAGCC	AATCTGCTTG	CGGTAGCTAG	TGCGATCTCC	TTGCAGAGTC	120
AGGCCATTGA	TATTGATGGA	GCCACTATAA	GGTGCCAAACA	GACCGATAAT	CTCATTGATC	180
GTGCTTGATT	TCCCAGCACC	ATTGAGACCA	ATCAAACCGA	CCAACTGCCC	ACTTTCAACA	240
GTAAGGACA	CATCTTTCAA	AACAGGAACA	TGAACATAGC	CACCTGTGAG	GTTTTTAATT	300
TCTAACATAT	TTTCTCCAAA	TCTGGTATAA	TGTAGCTATA	TTATATCAAA	ATTCAGTACA	360
GTAGAGGTAG	ATTTTATGTC	AGATTGCATT	TTTTGTAAAA	TCATCGCAGG	GGAAATTCCT	420
GCTTCGAAAG	TATATGAAGA	TGAGCAGGTC	CTTGCCCTTC	TTGATATCTC	TCAAGTAACA	480
CTAGGACACA	CCTTGGTCGT	GCCAAAGAA	CACTATCGCA	ATCTTTTGGA	GATGGATGCT	540
ACGAGCGCCA	CCAACTCTTT	GCCCAAGTAC	CAAAAGTAGC	TCAAAAAGTC	ATGAAAGTCA	600
CTAAGGCTGC	TGGTATGAAT	ATCATTTCOA	ACTGTGAAGA	AGTCGCTGGT	CAAACAGTTT	660
TTCATACTCA	CGTTACACCT	GTGCTCGCT	ACAGTGCTGA	CGATGACCTC	AAGATTGATT	720
TTATCGCCCA	CGAAACAGAC	TTTGAC				746

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 767 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

GATCCAAGCA	GTCCGTGATG	TAAGCTTTGA	AGTTAATGAA	GGAGAAGTTG	TTCCCTTAT	60
CGGTGCCAAC	GGTGAGGTA	AGACAATAAT	TCTTCGCACC	TTGTCAGGTT	TGGTTCGACC	120

WO 98/26072

PCT/US97/22578

-135-

```

AAGTTCAGGA AAGATTGAAT TTTTAGGTCA AGAATCCAA AAAATGCCAG CTCAGAAAAAT 180
TGTGGCAGGT GGTCTTTCAC AAGTTCAGGA AGGACGCCAC GTCCTTCCTG GCTTGACTGT 240
5  TATGGAAAAAT CTGAAATGG GAGCTTCTT AAAGAAAAAT CGTGAAGAAA ATCAAGCTAA 300
CTTGAAGAAG GTTTCTCAC GCTTTCCTCG TCTTGAAGAA CGTAAGAACC AAGATGCAGC 360
TACTCTTTCA GGAGGGGAAC AACAAATGCT TGCCATGGGA CGCGCTCTTA TGTCAACACC 420
10 AAAACTTCTT CTTTATAGATG AACCATCAAT GGGACTTGCC CCAATCTTCA TCCAAGAGAT 480
TTTTGATATC ATTCAAGATA TTCAGAAGCA AGGAACAACC GTCCTCTTGA TTGAACAAAA 540
15 TGCCAATAAA GCACCTTGCAA TCTCTGACCG AGGATATGTA CTGGAACAGG GAAATCGTCT 600
ATCAGGGACA GGGAAAGACT CGCTCATCAG AGGAGTCAGA GCATATCTAG GTGGTAAACA 660
TCCAGTGGAT TTTTGTCCGC AGTGAGTTCT GATCATCAT TTAGTTGGGG CTGTGTAGGT 720
20 TCAGTAAGTC GGTATCAAAA TCAGGGTGTG TTGCCGCAGT GGGGTCTG 767

```

(2) INFORMATION FOR SEQ ID NO:74:

```

25 (i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 695 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

```

```

30 (ii) MOLECULE TYPE: DNA (genomic)

```

(iii) HYPOTHETICAL: NO

```

35 (iv) ANTI-SENSE: NO

```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

```

GAGCTCGGTA CCCGGGGATC CTCAGAGTC GATAATCGT TGGTTGGACG AACCTCGAAA 60
CTGGAGCATG AGATTTCCTT TAGTTCGATC ATATCTTCCA TCGACAAGAA TGTCAATCAA 120
45 TGATAAGAGT TCCAGTTTAT CTGGAGTTTC CGGGATCAT TCTTCCCAAG TGTAGCCCGT 180
CCAGGACCAA ATGTCCTTGT CTGGCAATTC CTTTCGGATG CGTTTAACTA GAGGCAAGAG 240
50 AATGCCAGTA TTGAGAAAAG GCTCCCTCC CAGCAAGTC AAGCCTTGAA CATAGGGTTG 300
GGCAAGGTCT GCCATAATCT GCTCTTCTAA TTCTGCTGTA TAGGGAATGC CAGCATTAAA 360
55 AGACCAAGTC GCAACATTAT AACATCCCTC GCAGTGAAC ATACAGCCTG ATACATAGAG 420
AGAGTTGGCG ACGCCTTCGC GGTCCACAAA GTTAAAGGCC TTGTAGTCAA TGATACGACC 480
TTGACTAAGT TCTTCGCTT TCCATTCTTG TGGTTTTGGA TTATTCATTC GCTACCTCTA 540
60 TCCAATAACG CTCGACTCCA TTGGAGCAT CCTCAAAAT TCCACCATTG GCTAGAATGA 600

```

WO 98/26072

PCT/US97/22578

-136-

CTGCTCTGCT AGCAGGATTA TTCACGCTAC AGGGCACCAG AGCTTCTTG ATGTCCTTTC 660  
 CCTAGCAACT TCAAGCCCTG ACGGAAGTCT TTTT 695

5 (2) INFORMATION FOR SEQ ID No:75:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 723 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

15 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

20 (xi) SEQUENCE DESCRIPTION: SEQ ID No:75:

CTCGGTACCC GGGGCTCTC TAGAGTCGAC GGCTACAATG ATATTAAGAT GGATGATGTG 60  
 25 ATTGACGCGT ATGTCATGGA AGAAATCAAG AGATAAGATT TTTTGCTCCT TTCTTAGGTG 120  
 GTGAGGGACG CAAGCAAACC GATGGTTTCA TTGCTTATTT TTGAGCCTAG GGTCTCAAAA 180  
 30 ATCCCCCTGT ATGGGACTGA TAAATCAGTT CCATCACTTT CACCACGGCG AAGAAGCAGC 240  
 ATGACTTCAA ATTGAACTTC GTTTCAAATTT AAACGTAAAA TCAAGAAGTT TAAATAGCT 300  
 AGGTCTGCTG GCCTAGCTTT TGGTTCAAAG TAGAGAAAGG AATATCATGG TAAATCATTT 360  
 35 CCGTATAGAT CGTGTGGGCA TGGAAATCAA GCGTGAAGTC AATGAGATTT TGCAAAAAGAA 420  
 AGTCCGTGAT CCAGGTGTCC AAGGTGTGAC CATCACAGAT GTTCAGATGC TGGGTGACTT 480  
 40 GTCTGTTGCC AAGGTTTAT ACACCATTTT GAGTAACCTT GCTTCGGATA ACCAAAAAGC 540  
 CCAATCGGG CTTGAAAAAG CAACTGGTAC CATCAAAGT GAACCTTGGTC GCAATTGTAA 600  
 ATTGTACAAA TCCAGATTT GACCTTCGTC AAAGACGAGT CCATCGAGAT GGAACCAAGA 660  
 45 TTGACGAGAT GCTACGAAAT CTGGATAAGA CTAAGAAGA GGGGGTTGCC CCCCTTTTTT 720  
 GGG 723

50 (2) INFORMATION FOR SEQ ID No:76:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 970 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

55 (ii) MOLECULE TYPE: DNA (genomic)

60 (iii) HYPOTHETICAL: NO

WO 98/26072

PCT/US97/22578

-137-

(iv) ANTI-SENSE: NO

5

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:76:

	TGTCCTTATT TGCTGACCA AGTGCAAGCT GGTGGGATT GTGGTAACAT TGGATAAGAT	60
10	TTGACAAAGG AATTTCACAT ATGTAACGGT CTTACTCCAC GAAACGATTG ATATGCTTGA	120
	CGTAAAGCCT GAAGGTATCT ACGTTGATGC GACTTTGGGC GGAGCAGGAC ATAGCGAGTA	180
15	TTTATTAAGT AAATTAAGTG AAAAAGGCCA TCTCTATGCC TTTGACCAGG ATCAGAATGC	240
	CATTGACAAT GCGCAAAAC GCTTGGCACC TTACATTGAG AAGGGAATGG TGACCTTTAT	300
	CAAGGATAAC TTCCGTCATT TACAGGCACG TTTGGCGGAA GCTGGTGTTC AGGAJATTGA	360
20	TGGAATTTGT TATGACTTGG GAGTGTCTAG TCCTCAATTG GACCAGCGTG AGCGTGGTTT	420
	TTCTTATAAA AAGGATGCGC CACTGGACAT GCGGATGAAT CAGGATGCTA GTCTGACAGC	480
	CTATGAAGTG GTTAATCATT ATGACTATCA TGATTTGGTT CGTATTTTCT TCAAAACGG	540
25	TGAGGATAAA TTCTCTAAAC AGATTGCGCG TAAGATTGAG CAAGCGCTG AAGTGAAGCC	600
	GATTGAGACA ACGACTGAGT TAGCAGAGAT TATCAAGTTG CTCAAACCTG CCAAGGAAT	660
30	CAAGAAGAAG GGTCACTCTG CTAAGCAGAT TTCCAGGCT ATTCGAATTG AAGTCAATGA	720
	TGAACCTGGA GCGGCAGATG AGTCCATCCA GCAGGCTATG GATATGTTGG CTCTGGATGG	780
35	TAGAATTCA GTGATTACCT TTCAATCCTT AGAAGACCGC TTGACCAAGC AATTGTTCAA	840
	GGAGCTTCAA CAGTTGAAGT TCCAAAAGCG TTGCTTTTCA CCCAGATGAT CTCAGGCCCA	900
	AGATGGAATT GGTGTCCTCGT AAGCCAATCT TGCCAAGTGC GGAAGAGTTA GAAGCCAATA	960
40	ACCGTTGACT	970

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 954 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:77:

60	GAAAGAGTA ACTGATGCAC GTAACAGTAG GTGAATTAAT TGGTAATTTT ATTTTAATCA	60
----	--	----

WO 98/26072

PCT/US97/22578

-138-

	CTGGCTCTTT TATTCTTTTG CTAGTCTTGA TTAATAAATTT TGCATGGTCT AATATTACAG	120
5	GCATTTTCGA AGAAAGAGCT GAAAAAATTTG CTTCAGATAT TGACAGAGCT GAAGAAGCCC	180
	GTCAAAAAGC AGAAGTATTG GCTCAAAAAC GCGAAGATGA ATTGGCTGGT AGCCGTAAAG	240
	AAGCTAAGAC AATCATTTGAA AATGCAAAAGG AAACAGCTGA GCAAGTAAG GCTAATATCT	300
10	TAGCAGATGC TAAAGTAGAA GCGGGACACT TAAAGAAAA AGCCAATCAA GAAATTGCTC	360
	AAAAATAAGT AGAAGCTTTA CAGAGTGTTA AGGGTGAGGT CGCAGATTTG ACCATCAGCT	420
15	TAGCTGGTAA AATCATCTCA CAAAACCTTG ACAGTCATGC CCATAAAGCA CTCATTGATC	480
	AGTATATCGA TCAGCTAGGA GAAGCTTAAT GGACAAGAAA ACAGTAAAGG TAATTGAAAA	540
	ATACAGCATG CCTTTTGTCC AATTGGTACT TGAATAAGGA GAAGAAGACC GTATCTTTTC	600
20	AGACTTGACT CAAATCAAGC AAGTTGTGA AAAAAAGGTT CTGCTTCTT TTTTAAAAACA	660
	AGTGGCAGTA GACGAGTCGG ATAAGGAAAA AACAAATTGCT TTTTCCAAG ATTCTGTGTC	720
25	ACCTTTATTA CAAAACCTTTA TCCAGTTTCT GGCCTACAAT CACAGAGCAA ATCTTTTTTA	780
	TGATGTGCTT GTAGATTGCT TGAACCGACT TGAATAAGAA ACAATTCGAT TTGAAGTGAC	840
	GATTACGTCT GCTCATCCTC TAACTGATGA ACAGAAGACT CGTTTGCTCC CTTTGATTGA	900
30	GAAAAAATG TCTCTGAAAG TAAGGAGTGT AAAAGAACAA ATCGATGAAA GTCT	954

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1602 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

50	CCTGATTATA CCCAACCTCT TTGCATCAAG TCGGAAAAAT GAGTGAAATG GGTTTCCAGT	60
	TTTCTGAAA TAAGGTATCC TATATAAAGT ACCCTATGAT AACATGGAG GTATTGTGTA	120
55	TGGTTCAAAC AAGTCATTGA AGAAATACAA AACAAATGCC ACATTGTGGA AGTCATAGGA	180
	GATGTGATAT CTTACAAAAG GCAGGACGGA ACTATCTAGG GCTCTGTCTT TTTATGGTG	240
	AAAAAACACC ATCTTTCAGC GTTGTAGAGA ACAAGCAGTT TTACCACTGT TTTGGTTGTG	300
60	GTCGCTCAGG TGATGCTTTT AAAATTATATC GAGGAGTACC AAGGGGTTAC CTTTATGGAG	360



WO 98/26072

PCT/US97/22578

-139-

```

GCTGTCAAAA TCTTAGGTCA GCGTGTGGG ATTGAGGTTG AAAAACCGCT TTATAGTGAA 420
5 CAGAAGCCAG CCTCGCCTCA CCAAGCTCTT TATGATATGC ACGAAGATGC GGCTAAATTT 480
TACCATGCTA TTCTCATGAC AACGACTATG GCGGAAGAGG CCAGAAATTA CCTTTATCAG 540
CGGGGTTTGA CAGATGAAGT GCTTAAACAT TTTTGGATTG GTTTAGCACC TCCAGAACGA 600
10 AACATCTCT ATCAACGTTT GTCTGATCAG TATCGTGAAG AGGATTTACT GGATTACAGG 660
CTGTTTATC TTTCGGATGC CAATCAATTT GTAGACACCT TTCACAATCG CATTATTGTT 720
CCCCTGACAA ATGACCAAGG AAAGGTCATT GCCTTCTCAG GTCGTATCTG GCAAAAACG 780
15 GATTACAAA CTTCTAAGTA TAAAAACAGC CGTTCGACTG TAATTTTAA CAAAAGTTAC 840
GAATTATATC ATATGGATAG GGCAAAAGA TCTCTGGA AAGCTAGTGA GATTTACCTG 900
20 ATGGAAGGAT ATATGGATGT TATTGACGCC TATCGGGCTG GAATCGAAAA TGCTGTGGCG 960
TCGATGGGAA CGGCCTTGAG TCGAGAGCAT GTTGAGCATC TGAAAAGGTT AACCAAGAAA 1020
TTGGTTCCTG TTTACGATGG AGATAAGGCT GGGCAAGCCG CGACATTGAA AGCATTGGAT 1080
25 GAAATTGGTG ATATGCCTGT GCAAAATGTC AGCATGCTG ATAAC TTGGA TCCTGATGAA 1140
TATCTACAAA AAAATGCTCC ACAAGACTTG GCCTATCTAT TAACGAAAC TCGTATTAGT 1200
30 CCGATTGAGT TCTACATTCA TCAGTACAAA CCTGAAAACG GTGAAAATCT GCAGGCTCAG 1260
ATTGAGTTTC TTGAAAAAAT AGCTCCCTTG ATTGTTCAAG AAAAGTCCAT CGGTGCTCAA 1320
35 AACAGCTATA TTCATATTTT AGCTGACAGT CTGGCGTCTT TTGATTATAC CCAGATTGAG 1380
CAGATTGTTA ATGAGAGTCG TCAGGTGCAA AGGCAGAATC GCATGGAAAG AATTTCAGAA 1440
CGACGCCAA TCACCATGCC TGTCACCAAG CAGTATTCCG CTATTATGAG GGCAGAAGCC 1500
40 CATCTACTCT ATCGGATGAT GGAATCCCCT CTCGTTTGA ACGATTACCG TTTGCGAGAA 1560
GACTTTGCGT TTGCTACACC TGAATTTTCA GTCTTACATG AC 1602

```

(2) INFORMATION FOR SEQ ID NO:79:

- 45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7203 base pairs  
 (B) TYPE: nucleic acid  
 50 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 55 (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- 60 (x1) SEQUENCE DESCRIPTION: SEQ ID NO:79:

WO 98/26072

PCT/US97/22578

-140-

	CCTCCATCAA	ATCTGAGACT	GATTCAAAAAG	ACTGGCTCAT	ATTACGATTT	TGGTCTAAAT	60
5	GGGTTAACAC	TTGGAGCAAC	TTCCGATTTT	CGTCTAGTCT	AACATCAAAA	GGTAATCCCT	120
	GATATTGAAT	TGCCCTGACG	AAGGAAAATA	TTAATAGCTG	TTGTCAATAT	CATTCCCGAGA	180
	TTACTAAACA	CCTGTTGGGC	CTGCTCCTTA	ACCTCACTAT	CCAGACGGAT	GCTCATACTC	240
10	ATCTTTTGACA	TACTCTCACC	CTCTTTCCAT	AGACTATTTT	AACAAAAAAG	AAAGCTAATG	300
	TAAATCTATT	GGATATACGT	TAGCCTCTTC	TAATAGATTA	TTAAGCAATT	TTTAAAAACA	360
15	ACTCATCAAA	CAAACTCAAC	TGGTTATCCT	CTGGCATATT	TCCAAGAATA	CCCATCTCAT	420
	CCATCTTTTC	AACCAAGGTT	GATGAGAGTC	CACCACGCTT	GGGTAGTTCT	GTTTTAGAGA	480
	GGAAATCTCC	CTCTTCACGC	GCCCCGACCA	GTGCTTGGC	AAGCTTCTCT	CCCAGACCAT	540
20	CCATTGCTAC	AAATGTGGG	ATAAGGGTAT	CCCCGTCGAT	GAGGAACCTC	GTCCCTGCAC	600
	TACGGTAGAG	ATCTAATTG	CCAACTTGA	AACCTCGTTC	CCCATCTCA	TTGACAATCT	660
25	CAAGAGTTGT	ATAGAGATCG	ATTTCCACAT	TAGAGGCTTC	ATTGTTCTTC	CGTTTTTCAG	720
	AGATTCTTTC	CATTCTCGGC	TTGATGGCCT	CCAAGCCCCG	ACCCATGGTC	TTGATATCAA	780
	AAGCCTTAGC	ACGAATGGAG	AAGTAAGCAC	AGTAGTAATA	AATAGGATGG	TGAACCTTGA	840
30	AGTAAGCTAC	ACGCAAGGCC	ATCATACGAT	AGGCTGCCGC	ATGGGCCCTTA	GGGAACATGT	900
	ACTTAATTTT	CCCACAGGAT	TCGATATACC	ACTCTGGCAC	CTTATTAGCC	TTCATGGCTT	960
35	CGATATAGCC	ATTTCTCTCC	TCTTCTGAAA	TCTTTAGCCA	CAAAACCCCTA	CGTACCCGTT	1020
	CCATAATGGT	AAAGGCCATC	TTAGGTTCCA	GACCCGCATG	CATGAGGTAA	ACCATGATGT	1080
	CGTCCCGACA	ACCGATAACA	GTGATAGGT	CCGCTATTCC	TTGCTTAATC	AGATCTTGAG	1140
40	CATTCCCCAA	CCAAACGTCA	GTACCGTGGG	ACAGACCAGA	CAGCTGAAGC	AATTCTGGAA	1200
	AGGTTGTGCG	ATGGGTTTCG	TCTACCATTC	CACGTACGAA	ATTTGTTCCA	AACCTTGGAA	1260
45	TCCCTAACAT	ACCCGTAGCG	TTCCAATTTG	TTGAGGTGTT	ACCCCTAGCA	CATCATCCCC	1320
	AGAAAAGAGT	GCCATCACGC	CTTCGTCTAT	CATAGGAATT	TTATTAGGGT	CAATACCAGA	1380
	CAATCCTGA	AGTTTTTCGA	TCATGGTCGG	ATCATCATGT	CCAGTACAT	CGAGTTTGAG	1440
50	GACGTTCTCA	TCGATATCGT	GGAAATTAAT	GTGAGTGGTC	TGCCATTGAG	CCGTGACATC	1500
	ATCTGCTGGA	TACTGGACAG	GGCTAAATAT	GTAGACATCC	ATGTAGTTCC	GAATAACATC	1560
55	GATTCCCCCC	GGGTGTTGGC	CTGTTGTCCT	CTTGACACCC	GCCGCTCCTT	GAGCGAGGCG	1620
	TTCTACTTCT	GCATCAGCAT	AAAACCTTGC	ATAATCTCGC	TCGTAACCCCT	TGACAAATCC	1680
	ATAGGCAGTC	TTGGCAGCTA	CCGTACCAAC	TGTTCCCGCA	CGGAAGGCAT	ATTCCTCACCC	1740
60	AAAGATATCA	CGCACATCCA	AGTGGGCGCT	AGGCTGATCT	TCTCCCGAGA	AGTTCAAGTC	1800

WO 98/26072

PCT/US97/22578

-141-

	AATATCAGGA	ACCTTATCCC	CATCAAAACC	AAGGAAGGTC	TCAAACGGAA	TATCCTGTCC	1860
	GTTTTACTCG	AGTTTGTGAC	CACAGTTTGG	ACAGTCCCTA	TGGGGCATAT	CAAAATCCTGA	1920
5	ACCGTACGAA	CCACTGTGTA	TAAACTCACT	GTACTGCAC	TGACCACAGA	CATAGTGAGG	1980
	AGAGAGAGGA	TTGACCTCCG	TAATCCCAAT	CATGGTCGCA	ACGAACTAG	ATCCGACAGA	2040
10	CCCACGAGAA	CCAACCAAT	AACCCCGTTC	ATTAGAACGT	TGCACCAGCA	TCTGCGATGC	2100
	CAGATAAATC	ACAGCAAAATC	CATTCCCCAG	TATGGATGTT	AATTCTTTTT	CAATCCGCAA	2160
	ATCAACAATA	TCTGGCAGCG	GATTTCCATA	AATCTCAAAA	GCTTTCTTAT	AGGTCAAATC	2220
15	AGCAACTGTT	TCTTCAGCCT	TGTGATGAA	AGGCGTATAC	AAGTCACCCT	TAACTGCTTC	2280
	AACGGGTTC	AATAATTCTG	CCAAGGCATT	GGTGTTTTCA	ATAACCACTT	TACGAGCCAG	2340
20	TTCTCTTCCC	AAAAAGGCAA	ATTGATCCAA	CATCTCATT	GTCTGTCGAA	AATGAGCCTT	2400
	TGGAAGTGCT	GCTGCTGGG	CATGTTTACC	ATGACCATA	GTTGCGTTAA	TGATCGCACC	2460
	CTGTCCCAAA	CTACGGACGA	TAATTTTACG	ATAAATCTCT	TCTTCCGGTT	CGATATAGTG	2520
25	AACATTTCCC	GTAGCCAAAA	CAGGCTTGCC	AAGCGGCTCT	CCAACCTCTA	TCAAACCTTT	2580
	GATAATGGTC	TGGAGTTTCT	CCATATCCTT	GACCTGCTCT	TTAGCAATCA	AGGGCGCATA	2640
30	GATAGCCGGT	GGCATGACCT	CGATAAAGTC	ATAAATCTTG	GCCACCTCAA	CCGCGCGATC	2700
	CACACCTTGA	GAACGACCA	CGTCAAAAAC	TTCACCTCTT	GAACAGGCTG	AACCTAAAT	2760
	CAAGCCCTCT	CGATGGGCA	CTAGAACCGT	TCTCGGAATC	CGTGACACTC	CTTCAAAATA	2820
35	CTTGGTATTA	GACAAGGAAA	CCAGCTTAAA	GATATTTTTT	AGACCTACCT	GATTCTTGAC	2880
	ATAGATGGTC	GCATGCTTGA	TCCGAGCTTT	TTTGTAAAG	TCTGGACTGA	TTAGATCAAT	2940
40	GTGAGTCTA	GCTAAATCGG	TCACACCATG	TTTTTCTGCT	ACCTCTTTGA	TAAAGATAAA	3000
	GCAGACGACC	AGTCGCTTCC	GCATCGTAT	GGCCATGTGG	TGATGTTCCA	AGCCACACCA	3060
	AAACGCTTGG	TCAAAGGCCC	AAACCATGAT	TTATACTCAG	GATAGAGGTT	TCAGCAAACT	3120
45	CCAGGTATCA	ATAACGGCTG	ACTAATCTTT	GGCCATGACG	CTCATAAITA	GCATTATATA	3180
	AGCCAACGTC	AAAGGTAGCA	TTGTGGGCAA	CTAGGACCGT	ATCCTTGCAA	ATTCTTGGAA	3240
50	TTCTTGCAAA	ACTTGTCTTA	GTGGTTTGGC	ATTTTTTGACA	TGATCATCTG	TAAATCCAGT	3300
	TAACTCTGTA	GTAAGAGCTG	ACAAGGGATG	CCAGGATTTG	ATAAATTCAT	CAAATTCAGC	3360
	AATAACATTC	CCCTTTGTACA	TCTTAGAGGC	CGCAACCTGA	ATCAAGTCAT	TATAGATAGC	3420
55	TGAAAGTCCC	GTCTTTTCCA	CGTCAAAGAC	CACTAGGTTT	GCTTCTGATA	AGTCCATCTA	3480
	CATTCTGTTAT	AGACGATAGG	GACACGGTCC	TCCACGATAT	TGGCTTCTAT	CCCATAGATC	3540
60	AGCTGGATTC	CCGCTTTCTT	AGCCGCTCTA	TAGCCATGTG	GAAAGGACTG	GACATTCCCA	3600
	TGGTCTGTGA	TAGCAACCGC	CTTGTGTCCC	CACCTAGCAG	CTGTTGCAAC	AATCTCTTCG	3660

WO 98/26072

PCT/US97/22578

-142-

	ACCTCTGGCA AAGCATCCAT ACTCGACATG TTAGTATGAG CATGAAACTC AACCCGACGC	3720
5	TCACCTTCTG GCATCAAATC CTTCCGCTCA TAGTGAACAA CTTCTCGCAG ATCCTGTACG	3780
	TTCATAGTCA AATCGCGTGT GAAGTTATTC ATCTCCACAT TCCCTCGAAC TCGGAGCCAA	3840
	GAATTCCTCT TGATGAGGTC AAACCTTCTGG GCCTCTTCCT CGTTTCTAAC CCACCTTTTGC	3900
10	ATAGAAAAAC TTGAAGTATA GTCCGTCAAT TTAAGTTGA TTAAGAACAG ACCTGTTCTA	3960
	GTCACCTTTT GCTCCACATC AAAAACCAACC CCTTCAATA CCAGACGATT TTCCTCTGTC	4020
15	GTCACCTTGA TCATAGGAGT AATCTCCGCC TTATCCAGCT TGGGTTTAGC TGCAGCTTTT	4080
	TTTCGCTTGA AATCAAAGAC TGGTTTCTCT TCCGCTGGAG GAGGTGCCAT CTGCTCCAGT	4140
	TGTTCCATAG CACGGAGCGC TTCTCATTTG GCAGCTTGAA CAATCTGCTC ATTTTCAGCA	4200
20	TGAAAGGCCT CTTCCTGCTC TTGGGTGAG ACATCATCTT TCTCGACTTG ACAGTTAAAA	4260
	GTTCGAAAAA CAACTTTTTC AAGTTGTTTG GCTAAATTAG GAAGATGATT CTTCTTAAAA	4320
25	TGTTCTTAT CAATCGCTTC AGATCCTTCA ATAAATAGCT GATTACCCCT AGCACGAAT	4380
	TGCAAAATTT GATAAAGGGA CTTAAACCTT TGACTAGCAC ATGAGCCCTC AGAGAAAGCC	4440
	TCCCTATAGT AGGACTGCAA GAGCTGATTT GAAATTCCT GAGACCGAGC CTTAATTTCA	4500
30	AAAACAGCTT TATTGCCTGT CTTAGAAAAA TCTTCGCTCA AACCTTTCCT TAAATCTAAA	4560
	AAGATTTCAA TCGGTAAAAA ATTAGAAAAA ACGAAATGAA ACTCCATAC CTTACTAATT	4620
35	TTATGAACCA CAACTCGCTC AATATTGGCC TGTGCTAAAG CAGGAGCCTG TCTCATTTC	4680
	GCAGGCATCC CCAATTGATT CATCAAAAAT TCAAAACTAT TTGACATTCA TTTTCTCAC	4740
	ATTATTCTTC TACTATTTTA CCATATTTAG AGGTATTTTC TAAAGACAAA AGGAAGCCAC	4800
40	TAAATGACTT CCTTCTAGAG TGAGGACGGA TTAGTCTTCA CCTTTATTTT TCTTAATAAT	4860
	TTCTTCTTGT ACTGACTTAG GTACATCTTC GTAGTGGTCA AATACCATCA TGAATGTACC	4920
	ACGTCTTGA GATGCAGAAC GAAGAAGCTG TGCCTAACCG AACATTTTCA CAAGTGGAAC	4980
45	GTAAGCACGA ACGATTGGC TGTACCGTG TGCTTCCATA CCATCTACAC GTCCACGAGC	5040
	AGCAGTTACG TGACCCATAA CATCACCAG GTTTTCTTCT GGAACAGTGA TTGTTACAAG	5100
50	CATCATTGGT TCAAGGATAG CTGGTTGTGC TGATTTAGCA GCTTCTTTAA GGGAAAGTGA	5160
	AGCCGCAATC TTGAAGGCAG TTTGAGATGA GTCGACATCG TGATATGAAC CATCATAAAG	5220
55	CTTAGCTTTA ACGTCAACCA TTGGGTAAAC TGCAAGAACA CCGTTAGCCA TAGATTCTAC	5280
	CAAACCTTTT TCAACCGCTG GGATAAATTC ACGAGGAAC ACACCACCGA CGATTGCGTT	5340
	TTCGAATTCG AATCCTTTAC CTCTTCGTT TGGAGTAAAT TCAATCCATA CATCACCAG	5400
60	TTGACCTTTA CCACCAGACT GACGTTTGAA GAATCCGCTG GCTTGAGTAG AAGCGCGGAA	5460

WO 98/26072

PCI/US97/22578

-143-

	TGTTTCACGG TAAGATACTT GAGGAGCACC TACGTCGCT TCAACTTTGA ACTCAGCAGG	5520
	CATACGATCA ACAAGGACGT CAAGGTGAAG TTCACCCATA CCTGAGATAA CTGTTTCACC	5580
5	AGTTTCAACG TTTGTTTCAA CGCGGAATGT TGGATCTCT TCAGCCAAAT TTTGAAGGGC	5640
	GATACCCATC TTGCTTTGGT CAGCTTTAGA TTTTGGCTCA ACCATCAATT GGAATAACTGG	5700
10	TTCTGSAACG TTGATTGACT CAAGGATGAT TTTAGCTTTT TCATCTGTCA ATGAGTCACC	5760
	AGTTGTAGTA TCTTTCAAAC CAACGGCAGC AGCGATATCA CCTGAGTAAA CAGTGTCGAT	5820
	TTCTTGACGG CTGTTAGCGT GCATTTGAAG GATACGTCG ATACGTTTAC GTTTACCTTT	5880
15	AGAAGTATTC AATACGTATG AACCTGATTG AAGAACACCT GAGTAAACAC GGAAGAATGT	5940
	CAACAGACCT ACGAATGGGT CAGTCATGAT CTGGAAGGCA AGAGCTGCAA ATGGCTCTTC	6000
20	GTGAGATGCT GGACGAATTT CTTAGCGGTC TGTATCTGGG TTAATACCTT TGATTGCTGG	6060
	GATGTCAAGT GGAAGTTGAA GGTAGTCGAT AACCGCATCA AGCATCAATT GAACACCTTT	6120
	GTTTTGAAG GCTGAACACC ACAATACTGG GAAGAATTCA ACGTTGATAG TCGCTTTACG	6180
25	GATACAGCTT TCAATTCTTT CGTTAGTGAT TTCTTCACCT TCGAGGTATT TCATCATCAA	6240
	TTCTTCGTCA GTTCAGCAA CTGCTTCAAT CAATTTTCCA CGGTATTCCT GAGCTTGGTC	6300
30	AAGGATATCA GCTGGGATGT CTTCTTCAAG GATATCCGTA CCAAGGTCGT TAGTATAGAT	6360
	TTCAAGCTTC ATCTTGATCA AGTCAATGAT ACCACGGGAA TCACTCTCAG AACCGATTGG	6420
	CAATTTGATT GGGTGTGCAT TTGCTTGAAG ACGATCGTGA AGTGTGCTTA CAGAGTAAAG	6480
35	GAAATCAGCA CCGATTTTGT CCATTTTGTG GGCAAATACG ATACGTGGAA CTCCTGACTC	6540
	AGTTGCTTGA CGCCAAACTG TTTCAAGTTG AGGCTCAACA CCTGATTGTG AGTCAAGAAC	6600
40	GGTAACCGCA CCATCCAATA CACGAAGAGA ACGTTGTACT TCGATTGTGA AGTCCACGTG	6660
	TCCTGGTGTG TCGATGATGT TTACGCGGTG GTTGTTCAT TGAAGCTGTG TCGCAGCAGA	6720
	TGTGATCGTG ATACCACGTT CTGCTCTTG CTCATCCAG TCCATTGTG ACGCACCTTC	6780
45	GTGAGTTTCA CCGATTTTGT GATTTTACC AGTGTAGTAA AGAATACGCT CAGTAGTTGT	6840
	TGTTTATCCG GCATCGACGT GAGCCATGAT ACCGATATTA CGAGTTTTTT CAAGTAAAA	6900
50	TTGGCGTGCC ATGAGGTTTG TTTCTCTAT TTATTTTGA TTTCTATTCT ATTATAACAC	6960
	GATTTTAATA AAAACGGATA GGCAGGACCT ACCGTTCTCT AATGTTTCCA TGCTATTGTT	7020
	GGTTTCAACT TACGAGATGG TAAAGCTGAGT TATAGCTAAT ACTAATCGAT TTAGCTAATT	7080
55	TGAACCCGGG CTAAAGTTAG TTAGCCGATA TGAGCTGGAA CGGGATGCTG CGCGAAAAAG	7140
	ATAAACTCC TTGTATTCAT CGAATACTGC GTCAGTTTCC TATTTTCACC TTGCATCCTT	7200
60	ACC	7203

(2) INFORMATION FOR SEQ ID NO:80:

WO 98/26072

PCT/US97/22578

-144-

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1581 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

GCATACAAAG GGCATCAAGA ATATCCGTGT TGCCACAAGC TGCAGAAAG ATTTATGCCT	60
ACCGCCGTTA TGACCTTAAT GAATCTCCAA AGACCGCTTT AGACCTTATC ATCCCGATT	120
TGTTTTTGCA TATTTTGAAC CCTGCTGAAC GTGAAAGAAA ATTAAAGCGC GAAGGTGTAG	180
AAGAATTATA TCTCCTTGAT TTTAGTAGTC AATTCGCTAG TCTCACTGCA CAAGAATTCT	240
TTGCAACTTA TATCAAGGCT ATGAATGCCA AAATTATTGT TGCAGGTTTT GATTATACAT	300
TTGGTTCTGA CAAAAAACA GCAGAAGATT TAAAGGATTA CTTTGATGGA GAAGTTATCA	360
TTGTTCCACC TGTAGAAGAT GAGAAAGGAA AGATTAGTTC AACTCGTATC CGTCAAGCTA	420
TTTITAGATGG AAATGTGAAA GAAGCAGGAA AACTTTTGGG GGCACCGCTT CCATCAAGAG	480
GTATGGTAGT TCATGGTAAT GCTCGTGGTC GTACAAITGG TTATCCGACA GCGAATTTAG	540
TGCTTTTAGA CGTACTTAT ATGCCAGCAG ATGGCGTTTA TGTCGTTGAT GTTGAGATTC	600
AAAGCAGAA GTATCGTGCT ATGGCTAGTG TCGGGAAAAA TGTGACCTTT GATGGAAGA	660
AAGCAGTTT TGAAGTCAAT ATTTTGTATT TTAATCAAGA TATTTATGGG GAAACCGTCA	720
TGTTTATTG GCTTGATCGC ATTCGTGATA TGACCAAAAT TGACTCAGTT GACCAATTAG	780
TGGATCAGTT AAAGGCTGAT GAAGAAGTAA CTCGGAATTG GTCTTAAGAG CTTGAGTAAA	840
TAAAACAAAA AAGAGGTTGT CTGTAACCCA AAAGATAGAT GATTTAGTCT AACCTTTTGA	900
GTACGACAT TACCTCTTTT TATTCCTTTT CAAGGTGAA GCCTTCTCCT AGGATTTTCA	960
GGCGTCTGT AATAGTTATA AAGGCTTGAG GATCGATTCC ATGAATCAT TCCCTCGTTT	1020
TGCATATTTC ATTTCTCCG ACAATACAGT AGATGATTTT CAAATTTTCT TTACTATAGT	1080
AGCCTTGACC AGAATATAAA GTAACACCTC TTCCGAGGTC ATCATTAATC GCCTTAGCAA	1140
GTGTGTCAGC ACCTTTGTG ATAATCATAA AGCCTTTGCC GGCATATCCT CCTTACCAA	1200
TCAATCAAT AACACGAGAA ACAATAAAAT CAAACAAAG CGTGTAGGAA ACCAATCTCA	1260
AATCCTTGAA GATTAGGAGA ATCAACATGA GAATACAAA ATCTAAGATA AAGAGCAGTT	1320

WO 98/26072

PCT/US97/22578

-145-

	TTCCTATGGA TATATGAGTG TATTTGTTGA GAATACGAGC TAGAATATCA GTTCCGCCAG	1380
5	TTGTACCTCC ASCATTAAAA ATAATTCCAA GGCCAAATCC CAATAGGATT CCCGCTATAA	1440
	GGGCTGTGAT TAGTAAATCA CCTTGAAGAT CAATATGAAG GGGAAATATGC TCAAAAAAAG	1500
	CTAACCCAGGC GGACAAAGCT AAGGTTCCCTA GTAAACTAGA ATAGAGGGAT TTGGCTCCAA	1560
10	AGATCTTTCCA AGCTAGGATG A	1581

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 879 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:	
	CAAGTTTGTC GAATTGCCAA ACACAGTTGA AGGCTTGATT CACATCACTA ATCTACCTGA	60
	ATTTTATCAT TTCAATGAGC GTGATTGAC TCTTCGTGGA GAAAAATCAG GTATCACTTT	120
35	CCGAGTGGGT CAGCAGATCC GTATCCGTGT TGAAGAGCG GATAAATGA CTGGAGAGAT	180
	TGATTTTTC TCGTACCTA GTGAGTTTGA TGTGATTGAA AAAGGCTTGA AACAGCTAG	240
40	TCGTAGTGGC AGAGGGCGTG GTTCAAATCG TCGTTCGGAT AAGAAGGAAG ACAAGAGAAA	300
	ATCAGGAGCG TCAATGATA AGCGTAACAT TTCACAAAA GACAAGAGA AAAAAAGAAA	360
	GAAACCTTTT TACAAGGAAG TAGCTAAGAA AGGAGCCAAG CATGGCAAAG GGCAGGGGAA	420
45	AGGTCGTGCG ACAAATAAAA AAGGCACGCC ACGACTATAC AATCGTAGAT ACGCTAGAGG	480
	CAGGGATGGT CCTGACTGGA ACTGAAATCA AGAGTGACG AGCTGCTCGA ATTAATCTCA	540
50	AGGATGGCTT TGCTCAAGTG AAAAATGGAG AAGTTTGGCT GAGTAATGTT CATATCGCGC	600
	CTTACGAAGA GGGCAATATC TGGAAACGAG AACCAGAACG TCGTCGTAAA CTCCTGCTCC	660
	ATAAAAAAGA AATTCAAAAA TTGGAACAAG AGACCAAGG GACAGGAATG ACCTTAGTTC	720
55	CCCTTAAGGT CTATATAGAT GGCTACGCTA AGCTTCTTTT AGGACTTGCC AAGGGAAGCA	780
	TGACTATGAC AAACGGAGTC TATCAAACGT CGTGAGCAAA TCGAGATATC GCGCGTGTGA	840
60	TGAAGCTGTT AATCAGCGAT AAAGAGAGGA ATTGAGATG	879

(2) INFORMATION FOR SEQ ID NO:82:

WO 98/26072

PCT/US97/22578

-146-

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1550 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

20	AAAAGCTTAA TAAATCAATA ATTTCTTCTT TTATCCCCAA CCTGTGGATA AAGTTTGGTA	60
	ACATTGTGGA TTATTTTTC AAGCTTGTGG AAAATTCTTG CTATCTATGG TAAATATATCT	120
	CTAGATTAA AACTTTTAAAT AGTAAAGGAG GAGAAAGGAT TGAAAGAAAA ACAATTTTGG	180
25	AATCGTATAT TAGAATTTG CACAAGAAAGA CTGACTCGAT CCATGTATGA TTTCTATGCT	240
	ATTCAAGCTG AACTTATCAA GGTAGAGGAA AATGTTGCCA CTATATTTCT ACCTCGCTCT	300
	GAAATGGAAA TGGCTGGGA AAAACAATA AAGATATTA TTGTAGTAGC TGGTTTGAAG	360
30	ATTTATGAGC CTGAAATAAC TCCCACATAT ATTTTCACCA AACCTCAAGA TACGACTAGC	420
	TCACAAGTTG AAGAAGCTAC AAATTTAACT CTTTATGACT ATAGTCCAAA GTTAGTATCT	480
35	ATTCCTTATT CAGATACGGG ATTAAAAGAA AAGTATACCT TTGATAACTT TATTTCAAGG	540
	GATGGAATG TTTGGGCTGT ATCAGCCGCT TTAGCTGTCT CTGAAGATTT GGCTCTGACC	600
	TATAACCCCT TTTTATCTA TGGAGGACCA GGCCTTGGTA AGACTCACTT ATTAAACGCT	660
40	ATTGGAATG AAATTTCTAA AAATATTCCT AATGCGCGTG TTAATATATC CCCTGCCGAA	720
	AGCTTTATTA ATGACTTTCT TGATCACCTA AGACTTGGGG AAATGGAAAA GTTTAAAAAG	780
45	ACCTATCGTA GTCTTGATCT TTTGTTAATC GATGATATCC AGTCACCTAG CGGAAAAAAA	840
	GTGCAACTC AGGAAGAATT TTTCAATACC TTTAACGCCC TTCTAGACAA GCAAAAACAG	900
	ATTGTCCTAA CGAGTGATCG TAGTCCAAAA CATCTAGAAG GGCTCGAGGA GAGGCTTGTC	960
50	ACGCGTTTAA GTTGGGGATT GACACAACT ATCACACCCC CTGACTTTGA AACACGTATT	1020
	GCCATTTTAC AAGTAAAAAC GGAACATTTA GGCTACAATT TCCAAAGTGA TACTCTAGAA	1080
55	TACCTAGCTG GGCATTTTGA TTCAATGTTT CGAGATCTTG AGGAGCCCAT CAACGACATC	1140
	ACTTTAATTG CCAGAGTAAA AAAAATCAAG GATATCACTA TTGATATTGC TGCAGAAAGCC	1200
	ATTAGAGCCC GCAACAAGA TGTTAGCCAA ATGCTCGTCA TCCCAATTGA TAAATCCAA	1260
60	ACTGAAGTTG GTAACCTTTA TGGTCTTAGT ATCAAAGAAA TGAAGGGAAG TAGACGCGCTT	1320



WO 98/26072

PCT/US97/22578

-147-

CAAAAATATTG TTTTGCCCCG TCAAGTAGCC ATGTATTAT CTAGAGAACT AACAGATAAT 1380  
 AGCTCTCCAA AAATTGGGAA GGAATTGGGG GAAAAGTCAT ACCACAGTCA TTCATGCCCA 1440  
 5 TGCCAAATAA AATCTTGAA TTGATCAAGA CGATAATTGA CGTTAGAAA TTGAATCATC 1500  
 AAAAGGAAAA TCAATAAATT TGTGAAACT TTAGGTTTTC ACCTTTTAGC 1550  
 10 (2) INFORMATION FOR SEQ ID NO:83:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1292 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: DNA (genomic)  
 20 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:  
 GGTATGCGCC AAACTTCTT ATCAAAAAGA ATCTAGCCAA AGAAGCGACT GCTCAAGCTG 60  
 30 TAGGTGAAC TCGTGATAA CAAAAATCGG AAGAAAAAGC TCACGCTGAG ATGATTGCAG 120  
 AAGGAAAAAG AATTAAAGCA CAACCTGAAG CAGAAGAAAC TGTGTAGAA TTTGTTGAAA 180  
 35 AAGTTGGTCC AGATGGTCGT ACCTTTGGTT CTATTACCAA TAAGAAGATT GCAGAAGAAT 240  
 TGCAAAAGCA ATTTGGAATT AAGATTGATA AACGTCATAT TCAAGTACAA GCTCCGATTC 300  
 GAGCGGTTGG TTGATTGAT GTGCCAGTGA AAATCTATCA AGATATCACA AGTGTAAATCA 360  
 40 ATCTTCGTGT GAAAGAAGGA TAAGTTTACA CCTTCTTGAC AAGATTCTAA AAGGAAGGGA 420  
 AGTCTGATGG CAGAAGTAGA AGAGTTACGA GTACAACCTC AAGATATCTT AGCTGAGCAA 480  
 45 TCCGTTTATG GGGCTATCTT TATTGATGAG AGTAACTTG TTTTGTGCG AGAATACATT 540  
 GAGTCTCGGG ACTTTTTTAA GTATGCCCAT CGTTTGATTI TCCAAGCCAT GGTGCGATTTA 600  
 TCCGATCGTG GTGATGCCAT AGATGCAACA ACGGTTGCTA CTATCCTTGA TAATCAAGGT 660  
 50 GATTACAGA ATATTGGTGG CTTGTCTTAC TTGGTTGAGA TTGTTAATCT TGTGCCAACT 720  
 TCTGCTAATG CGAGTATTA TGCTAAGATT GTTGCAAGAA AAGCAATGCT ACGTCGTTAA 780  
 55 ATTGCCAAGT TGACAGAGTC TGTCAACCAA GCTTACGAA GGTCAACACC AGCTGATGAA 840  
 ATTATTGCTC AGGCAGAAAA AGGGTTGATT GATGTCAGTG AAAATGCAAA TCGAAGCGGG 900  
 TTTAAGAACA TTCGAGATGT GTTGAATCTC AACTTTGGAA ATCTGGAAGC TCGCTCGCAA 960  
 60 CAAACGACCG ATATTACAGG TATTGCGACA GGTTATCGTG ATTTGGATCA TATGACAACA 1020

WO 98/26072

PCT/US97/22578

-148-

5 GGACTTCATG AGGAGGAGTT GATTATCTTA GCAGCTCGTC CAGCAGTTGG TAAGACAGCA 1080  
 TTTGCCTTGA ATATCGCTCA GAATATTGGG ACTAAGTTGG ACAAACGGT TGCTATTTTT 1140  
 10 TCACCTCGAAA TGGGTGCGGA AAGCTTGGTA GACCGTATGT TAGCTGCAGA AGGCTTGGTG 1200  
 GAGTCACATT CTATCCGTAC AGGGCAATTG ACAGATGAGG AGTGGCAAAA ATATACTATT 1260  
 15 GCTCAGGGTA ATCGTACTAA CGCCAGTATC TA 1292

(2) INFORMATION FOR SEQ ID NO:84:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1876 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

25 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

30 AGGCTGATCC TGGCGTTATT GACCCATGGA AGAGCACCTA GAAGGAGATC ATTCCCAGAC 60  
 GATATTTTAG TTTTATCCTA GTAGCCTTCC CTGGCTATTT TAGGAGCTCG TCTCTACTAT 120  
 35 GTATTTCCGA TTTGATTACT ATAGTCAGAA TTTAGGAGAG ATTTTGGCCA TTGGAATGGT 180  
 GGTGGCCAT TTACGGTGGT TGATAACTGG GGCCTTTGTG CTCTATAICT TTGCTGACCG 240  
 TAAACTCATC AATACTTGGG ATTTTCTAGA TATTGCGGCG CCTAGCGTTA TGATTGCTCA 300  
 40 AAGTTTGGGG CGTGGGGTA ATTTCTTTAA CCAAGAACTT TATGGTGCAA CAGTGGATAA 360  
 TCTGGATTAT CTACCTGGCT TTATCCGTGA CCAGATGTAT ATTGAGGGGA GCTACCGTCA 420  
 45 ACCGACTTTC CTTTATGAGT CTCTATGGAA TCTGCTGGC TTTGCCITGA TTCTGATTTT 480  
 TAGACGGAAA TGGAAAGATC TCAGACGAGG TCATATCACG GCCTTTTACT TGATTGGTA 540  
 50 TGGTTTCGGT CGTATGGTCA TCGAAGGTAT GCGAACAGAT AGTCTCATGT TCTTCGGACT 600  
 TCGAGTGTCC CAATGGCTGT CAGTTGTCCT TATCGSTCTC GGTATAATGA TCGTTATTTA 660  
 55 TGCATATATT CTGTGTGCC TAGCTTTGAT TGTCTTTTGT GTCTATCTGA TCATTACTGT 780  
 ACAAAGCCTT GGTCTGTGTA TCGATGAAAC AGAAAAGACG ATTAACACCT TGACTTCAGA 840  
 TGTGGATGTG ACCTTGCACT ACACCAATGA GTTGTGGCT AAGGTCAATG TCTTGGCAGA 900  
 60 TGATATCAAT GTCAAGGTGG CTACGATTGA TCCACTCTTC AGTGCTGTTG CAGATTTATC 960

WO 98/26072

PCT/US97/22578

-149-

```

TCTATCTGTT TCAGACCTCA ATGACCATGC GCGTGTCTTG AGCAAGAAAG CTTCATCAGC 1020
5 TGGTTCAAAA ACATCAAGA CTGGTGCAAG TCTGTCAGCT CTTGTCCTTG CAAGTAAATT 1080
TTTCAAAAA TAAAAAGGA GAATCCTTAT GGGTAAATTA TCCTCAATCC TTTTAGGAAC 1140
GTTTCAGGT GCAGCTCTTG CCTTGTTTT AACAAAGTAT AAGGGCAAC AAGTTTGAG 1200
10 TCAGGCTCAA GATTTTCTAG ATGATTGAG AGAAGATCCG GAGTATGCCA AGGAGCAAGT 1260
CTGTGAAAA CTGACAGAAG TTAAGGAGCA GGCTACAGAT TTTGTTCTGA AAAAAAGA 1320
ACAGGTTGAG TCAGGTGAAA TCACTGTGGA CAGTACTACT GCTCAAGCTA AATCCTATGC 1380
15 TTTCAAGCG ACAGAAAGCAT CAAAAATCA ATTAAATAAT CTCAGGAAC AATGGCAAGA 1440
AAAAGCCGAA GCTCTTGATG ACTCAGAAGA GATTGTGATT CATATAACAG AAGAATAAAC 1500
20 CATCACCATC TCCGGACGGA CTATGTATCT GGGGATGGTG ATTTTATCTT GGAATCTAGT 1560
CTTTGTGGTA TAATAATTAC TATGCAGAAA AAACCAAGT CAGCCTATGT GCATATCCCA 1620
TTTGTACCC AGATTGTGTA TTATTGTGAT TTTTCAAGG TCTTCATCAA AAATCAGCCA 1680
25 GTGCAGAGCT ATTTAGAGCA TCTGCTGGA GAGTTTCGTT CTTATGATAT TGAAGGTTG 1740
TCAACCCCTT ATATCGGTGG TGGAACACGA CAGCCCTGTC GGCTCCGCAA CTGGAGGTGT 1800
30 TACTGAATGG CTGACTAAA AACTTGGATT TGCTGCTTG GAGAGTGACC ATTGAAGCCA 1860
TCCAGCGCAT TTGAA 1876

```

```

(2) INFORMATION FOR SEQ ID NO:85:
35 (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 1574 base pairs
      (B) TYPE: nucleic acid
      (C) STRANDEDNESS: single
      (D) TOPOLOGY: linear
40 (ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: NO
45 (iv) ANTI-SENSE: NO
50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

```

```

TTGGAAGATT TCCCACTTTC AGTGACCAAC CCATACGGTC GTACTAAGCT CATGCTAGAG 60
55 GAAATTTTGA CTGATATTTA CAAAGCAGAC TCAGAAATGGA ATGTTGTCTT GCTTCGTAC 120
TTTAACCCAA TCGGAGTCCA TGAGAGTGGT GATTGCGAG AAAATCCAAA CGGTATTCCA 180
AACAACTCTT TGCCATATGT GACTCAAGTA GCCGTTGGAA AATTAGAGCA AGTGCAAGTG 240
60 TTTGAGACG ATTACGATAC GGAAGATGGA ACAGGTGTTG GTGACTATAT CCACGTTGTC 300

```

WO 98/26072

PCT/US97/22578

-150-

GATTTGGCTA AGGGTCACGT TGCAGCTTGG AAAAAATCC AAAAAGGTT AGGACTAAAC 360  
 GTTTATAACC TTGGAACG TAAAGGTTAC TCAGTCTTGG AAATTATCCA AAACATGGAA 420  
 5 AAAGCGGTGG GATGTCTAT TCCTTACCGC ATCGTAGAAC GTCGCCAGG TGATATCGCT 480  
 GCCTGCTACT CAGACCCAGC AAAGGCTAAA GCAGAACTCG GTTGGGAAGC AGAACTCGAC 540  
 10 ATCACCCAAA TGTTGTGAAG CCATGGCGTT GGCAGAGCAA GCATCCAAAT GGATTTGAAG 600  
 ACTAAGATGA TGATTTCAT CATCGTCCCT TGTTTAACGA AGAGGAAGTA CTTCCTCTTT 660  
 15 TTTATCAGGC TCTGGAAGCT TTACTTCCAG ATTTGGAAAC AAAATCAGAT ATGTCITTTGT 720  
 CGATGATGGA TCAAGTATG GACCTTGGAA ACTCTTAAAG GCCTATCGGG AGCAAAATCC 780  
 GGCAGTCCAT TATATTTCTT TCTCTCGAAA TTTTGGCAAA GAAGCAGCCC TTTATGCAGG 840  
 20 CTTCGAATAT GCGACAGGAG ATTTGGTGGT GGTGATGGAT GCAGACCTCC AAGATCCTCC 900  
 TAGTATGTTG TTTGAGATGA AAAATGTACT AGACAAAAAT GTAGACTTGG ACTGCGTTGG 960  
 GACACGGAGA ACTAGTCGGG AGGAGAAACC CTTCTTTGCG AGTTTCTGTG CTGTTCTCTT 1020  
 25 TTATCGCCTC ATGCAAAAAA TCAGCCAGT AGCTCTGCCG TCGGGTGTCC GTGATTTTGG 1080  
 TATGATGAGA AGGCTGTGG TCGATGCCAT TTTAAGCTTG ACTGAGTCCA ATCGTTTTTC 1140  
 30 TAAGGGACTC TTGCTCGGG TCGGCTTTAA AACCACATAT CTGGACTATC CAAATGTGCA 1200  
 AAGGCAGGCT GGCAAGACCA GTTGGAGTTT TAGGCAACTT TTTTTTACT CCATTGAAGG 1260  
 GATTGTAAAT TTTTCAGATT TCCCTTTGAC TATAGCCTTT GTAGCTGGTC TCCTATCTTG 1320  
 35 TTTTCTTCT CTGCTGATGA CTTTTTTTGT TGTGGTTCGG ACCCTCATTT TGGGCAATCC 1380  
 GACATCTGGT TGGACCTCTC TGATGGCTGT TATTCTCTAT CTGAGAGGCA TTCAACTCTT 1440  
 40 GACCATTGGG ATTCCTGGTA AGTATAATCA GTAAGATTAT TTAGAGACTA AAAAAAGAC 1500  
 CACTTTATCT TATCAAGAA AAAGTGACCT TCCTGATTTT ACAGAAACCT AAAGTGAAAA 1560  
 GACTATAATT TTCC 1574

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 969 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

60



WO 98/26072

PCT/US97/22578

-152-

	1	5	10	15	
	GAG CAA TCC GTT TTA GGG GCT ATC TTT ATT GAT GAG AGT AAA CTT GTT				96
	Glu Gln Ser Val Leu Gly Ala Ile Phe Ile Asp Glu Ser Lys Leu Val				
5	20		25	30	
	TTT GTG CGA GAA TAC ATT GAG TCT CGG GAC TTT TTT AAG TAT GCC CAT				144
	Phe Val Arg Glu Tyr Ile Glu Ser Arg Asp Phe Phe Lys Tyr Ala His				
	35		40	45	
10					
	CGT TTG ATT TTC CAA GCC ATG GTC GAT TTA TCC AAT CGT GGT GAT GCC				192
	Arg Leu Ile Phe Gln Ala Met Val Asp Leu Ser Asp Arg Gly Asp Ala				
	50		55	60	
15					
	ATA GAT GCA ACA ACG GTT CGT ACT ATC CTT GAT AAT CAA GGT GAT TTA				240
	Ile Asp Ala Thr Thr Val Arg Thr Ile Leu Asp Asn Gln Gly Asp Leu				
	65		70	75	
20					
	CAG AAT ATT GGT GGC TTG TCT TAC TTG GTT GAG ATT GTT AAT TCT GTG				288
	Gln Asn Ile Gly Gly Leu Ser Tyr Leu Val Glu Ile Val Asn Ser Val				
	85		90	95	
25					
	CCA ACT TCT GCT AAT GCG GAG TAT TAT GCT AAG ATT GTT GCA GAA AAA				336
	Pro Thr Ser Ala Asn Ala Glu Tyr Tyr Ala Lys Ile Val Ala Glu Lys				
	100		105	110	
30					
	GCA ATG CTA CGT CGT TTA ATT GCC AAG TTG ACA GAG TCT GTC AAC CAA				384
	Ala Met Leu Arg Arg Leu Ile Ala Lys Leu Thr Glu Ser Val Asn Gln				
	115		120	125	
35					
	GCT TAC GAA GCG TCA CAA CCA GCT GAT GAA ATT ATT GCT CAG GCA GAA				432
	Ala Tyr Glu Ala Ser Gln Pro Ala Asp Glu Ile Ile Ala Gln Ala Glu				
	130		135	140	
40					
	AAA GGG TTG ATT GAT GTC AGT GAA AAT GCA AAT CGA AGC GGG TTT AAG				480
	Lys Gly Leu Ile Asp Val Ser Glu Asn Ala Asn Arg Ser Gly Phe Lys				
	145		150	155	
45					
	AAC ATT CGA GAT GTG TTG AAT CTC AAC TTT GGA AAT CTG GAA GCT CGC				528
	Asn Ile Arg Asp Val Leu Asn Leu Asn Phe Gly Asn Leu Glu Ala Arg				
	165		170	175	
50					
	TCG CAA CAA ACG ACC GAT ATT ACA GGT ATT GCG ACA GGT TAT CGT GAT				576
	Ser Gln Gln Thr Thr Asp Ile Thr Gly Ile Ala Thr Gly Tyr Arg Asp				
	180		185	190	
55					
	TTG GAT CAT ATG ACA ACA GGA CTT CAT GAG GAG GAG TTG ATT ATC TTA				624
	Leu Asp His Met Thr Thr Gly Leu His Glu Glu Glu Leu Ile Ile Leu				
	195		200	205	
60					
	GCA GCT CGT CCA GCA GTT GGT AAG ACA GCA TTT GCC TTG AAT ATC GCT				672
	Ala Ala Arg Pro Ala Val Gly Lys Thr Ala Phe Ala Leu Asn Ile Ala				
	210		215	220	
65					
	CAG AAT ATT GGG ACT AAG TTG GAC AAA ACG GTT GCT ATT TTT TCA CTC				720
	Gln Asn Ile Gly Thr Lys Leu Asp Lys Thr Val Ala Ile Phe Ser Leu				
	225		230	235	
70					
	GAA ATG GGT GCG GAA AGC TTG GTA GAC CGT ATG TTA GCT GCA GAA GGC				768
	Glu Met Gly Ala Glu Ser Leu Val Asp Arg Met Leu Ala Ala Glu Gly				
	245		250	255	

WO 98/26072

PCT/US97/22578

-153-

5	TTG GTG GAG TCA CAT TCT ATC CGT ACA GGG CAA TTG ACA GAT GAG GAG Leu Val Glu Ser His Ser Ile Arg Thr Gly Gln Leu Thr Asp Glu Glu	816
	260 265 270	
	TGG CAA AAA TAT ACT ATT GCT CAG GGT AAT CTA GCT AAC GCC AGT ATC Trp Gln Lys Tyr Thr Ile Ala Gln Gly Asn Leu Ala Asn Ala Ser Ile	864
	275 280 285	
10	TAT ATC GAT GAT ACG CCA GGT ATT CGG ATT ACA GAG ATT CGT TCT CGT Tyr Ile Asp Asp Thr Pro Gly Ile Arg Ile Thr Glu Ile Arg Ser Arg	912
	290 295 300	
15	TCT CGT AAA TTG GCT CAA GAA ACT GGA AAT CTT GGT TTG ATT TTG ATA Ser Arg Lys Leu Ala Gln Glu Thr Gly Asn Leu Gly Leu Ile Leu Ile	960
	305 310 315 320	
20	GAC TAT TTG CAA CTT ATC ACG GGA ACT GGT CGA GAA AAT CGT CAA CAA Asp Tyr Leu Gln Leu Ile Thr Gly Thr Gly Arg Glu Asn Arg Gln Gln	1008
	325 330 335	
25	GAA GTT TCT GAA ATT TCT CGT CAG TTG AAA ATA CTA GCC AAG GAA TTG Glu Val Ser Glu Ile Ser Arg Gln Leu Lys Ile Leu Ala Lys Glu Leu	1056
	340 345 350	
30	AAG GTT CCA GTA ATC GCT CTG AGT CAG CTT TCT CGT GGT GTA GAA CAA Lys Val Pro Val Ile Ala Leu Ser Gln Leu Ser Arg Gly Val Glu Gln	1104
	355 360 365	
35	CGT CAG GAC AAG AGA CCG GTC TTG TCT GAT ATT CGT GAA TCT GGG TCT Arg Gln Asp Lys Arg Pro Val Leu Ser Asp Ile Arg Glu Ser Gly Ser	1152
	370 375 380	
40	ATT GAG CAG GAC GCT GAT ATC GTA GCT TTT CTC TAT CGC GAT GAC TAC Ile Glu Gln Asp Ala Asp Ile Val Ala Phe Leu Tyr Arg Asp Asp Tyr	1200
	385 390 395 400	
45	TAT GAA CGT GGT GGT GAA GAA GAG GAG GGT ATC CCA AAT AAT AAG GTG Tyr Glu Arg Gly Gly Glu Glu Glu Glu Gly Ile Pro Asn Asn Lys Val	1248
	405 410 415	
50	GAA GTT ATT ATC GAG AAA AAC CGT AGT GGA GCT CGT GGA ACA GTG GAA Glu Val Ile Ile Glu Lys Asn Arg Ser Gly Ala Arg Gly Thr Val Glu	1296
	420 425 430	
55	TTG ATT TTC CAA AAA GAA TAC AAT AAA TTT TCA AGT ATC TCA AAG AGG Leu Ile Phe Gln Lys Glu Tyr Asn Lys Phe Ser Ser Ile Ser Lys Arg	1344
	435 440 445	
60	GAG GCA TAA Glu Ala	1353
	450	
55	(2) INFORMATION FOR SEQ ID NO:88:	
	(1) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 450 amino acids	
	(B) TYPE: amino acid	
60	(D) TOPOLOGY: linear	





WO 98/26072

PCT/US97/22578

- 155 -

```

Ser Arg Lys Leu Ala Gln Glu Thr Gly Asn Leu Glu Ile Leu Ile Leu Ile Leu Ile
305          310          315
5 Asp Tyr Leu Gln Leu Ile Thr Gly Thr Gly Arg Glu Asn Arg Gln Gln
    325          330          335
Glu Val Ser Glu Ile Ser Arg Gln Leu Lys Ile Leu Ala Lys Glu Leu
340          345          350
10 Lys Val Pro Val Ile Ala Leu Ser Gln Leu Ser Arg Gly Val Glu Gln
    355          360          365
Arg Gln Asp Lys Arg Pro Val Leu Ser Asp Ile Arg Glu Ser Gly Ser
370          375          380
15 Ile Glu Gln Asp Ala Asp Ile Val Ala Phe Leu Tyr Arg Asp Asp Tyr
    385          390          395
Tyr Glu Arg Gly Gly Glu Glu Glu Gly Ile Pro Asn Asn Lys Val
400          405          410          415
Glu Val Ile Ile Glu Lys Asn Arg Ser Gly Ala Arg Gly Thr Val Glu
420          425          430          435
25 Leu Ile Phe Gln Lys Glu Tyr Asn Lys Phe Ser Ser Ile Ser Lys Arg
    440          445
Glu Ala
450
30 (2) INFORMATION FOR SEQ ID NO:89:
    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1785 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear
    (ii) MOLECULE TYPE: DNA (genomic)
    (iii) HYPOTHETICAL: NO
    (iv) ANTI-SENSE: NO
    (ix) FEATURE:
        (A) NAME/KEY: CDS
        (B) LOCATION: 1..1782
        (D) OTHER INFORMATION: DNA G
50
    (xi) SEQUENCE DESCRIPTION: SEQ ID No:89:
ATG ATA ACC ATG GAG GTA TTG TGT ATG GTT GAC AAA CAA GTC ATT GAA
Met Ile Thr Met Glu Val Leu Cys Met Val Asp Lys Gln Val Ile Glu
1      5      10      15
GAA ATC AAA AAC AAT GCC AAC ATT GTG GAA GTC ATA GGA GAT GTG ATT
Glu Ile Lys Asn Asn Ala Asn Ile Val Glu Val Ile Gly Asp Val Ile
20      25      30

```



WO 98/26072

PCT/US97/22578

-157-

	Phe Met Asp Val Ile Ala Ala Tyr Arg Ala Gly Ile Glu Asn Ala Val	
	275 280 285	
5	GCG TCG ATG GGA ACG GCC TTG ACT CGA GAG CAT GTT GAG CAT CTG AAA Ala Ser Met Gly Thr Ala Leu Ser Arg Glu His Val Glu His Leu Lys	912
	290 295 300	
10	AGG TTA ACC AAG AAA TTG GTT CTT GTT TAC GAT GGA GAT AAG GCT GGG Arg Leu Thr Lys Lys Leu Val Leu Val Tyr Asp Gly Asp Lys Ala Gly	960
	305 310 315 320	
15	CAA GCC GCG ACA TTG AAA GCA TTG GAT GAA ATT GGT GAT ATG CCT GTG Gln Ala Ala Thr Leu Lys Ala Leu Asp Glu Ile Gly Asp Met Pro Val	1008
	325 330 335	
20	CAA ATC GTC AGC ATG CCT GAT AAC TTG GAT CCT GAT GAA TAT CTA CAA Gln Ile Val Ser Met Pro Asp Asn Leu Asp Pro Asp Glu Tyr Leu Gln	1056
	340 345 350	
25	AAA AAT GGT CCA GAA GAC TTG GCC TAT CTA TTA ACG AAA ACT CGT ATT Lys Asn Gly Pro Glu Asp Leu Ala Tyr Leu Leu Thr Lys Thr Arg Ile	1104
	355 360 365	
30	AGT CCG ATT GAG TTC TAC ATT CAT CAG TAC AAA CCT GAA AAC GGT GAA Ser Pro Ile Glu Phe Tyr Ile His Gln Tyr Lys Pro Glu Asn Gly Glu	1152
	370 375 380	
35	AAT CTG CAG GCT CAG ATT GAG TTT CTT GAA AAA ATA GCT CCC TTG ATT Asn Leu Gln Ala Gln Ile Glu Phe Leu Glu Lys Ile Ala Pro Leu Ile	1200
	385 390 395 400	
40	GTT CAA GAA AAG TCC ATC GCT GCT CAA AAC AGC TAT ATT CAT ATT TTA Val Gln Glu Lys Ser Ile Ala Ala Gln Asn Ser Tyr Ile His Ile Leu	1248
	405 410 415	
45	GCT GAC AGT CTG GCG TCC TTT GAT TAT ACC CAG ATT GAG CAG ATT GTT Ala Asp Ser Leu Ala Ser Phe Asp Tyr Thr Gln Ile Glu Gln Ile Val	1296
	420 425 430	
50	AAT GAG AGT CGT CAG GTG CAA AGG CAG AAT CGC ATG GAA AGA ATT TCC Asn Glu Ser Arg Gln Val Gln Arg Gln Asn Arg Met Glu Arg Ile Ser	1344
	435 440 445	
55	AGA CCG ACG CCA ATC ACC ATG CCT GTC ACC AAG CAG TTA TCG GCT ATT Arg Pro Thr Pro Ile Thr Met Pro Val Thr Lys Gln Leu Ser Ala Ile	1392
	450 455 460	
60	ATG AGG GCA GAA GCC CAT CTA CTC TAT CGG ATG ATG GAA TCC CCT CTT Met Arg Ala Glu Ala His Leu Leu Tyr Arg Met Met Glu Ser Pro Leu	1440
	465 470 475 480	
65	GTT TTG AAC GAT TAC CGT TTG CGA GAA GAC TTT GCA TTT GCT ACA CCT Val Leu Asn Asp Tyr Arg Leu Arg Glu Asp Phe Ala Phe Ala Thr Pro	1488
	485 490 495	
70	GAA TTT CAG GTC TTA CAT GAC TTG CTT GGC CAG TAT GGA AAT CTT CCT Glu Phe Gln Val Leu His Asp Leu Leu Gly Gln Tyr Gly Asn Leu Pro	1536
	500 505 510	
75	CCA GAA GTT TTA GCA GAG CAG ACA GAG GAA GTT GAA AGA GCT TGG TAC Pro Glu Val Leu Ala Glu Gln Thr Glu Glu Val Glu Arg Ala Trp Tyr	1584

WO 98/26072

PCT/US97/22578

-158-

	515	520	525	
5	CAA GTT TTA GCT CAG GAT TTG CCT GCT GAG ATA TCG CCG CAG GAA CTT Gln Val Leu Ala Gln Asp Leu Pro Ala Glu Ile Ser Pro Gln Glu Leu 530 535 540			1632
10	AGT GAA GTA GAG ATG ACT CGA AAC AAG GCT CTC TTG AAT CAG GAC AAT Ser Glu Val Glu Met Thr Arg Asn Lys Ala Leu Leu Asn Gln Asp Asn 545 550 555 560			1680
15	ATG AGA ATC AAA AAG AAG GTG CAG GAA GCT AGC CAT GTA GGA GAT ACA Met Arg Ile Lys Lys Lys Val Gln Glu Ala Ser His Val Gly Asp Thr 565 570 575			1728
20	GAT ACA GCC CTA GAA GAA TTG GAA CGT TTA ATT TCC CAA AAG AGA AGA Asp Thr Ala Leu Glu Glu Leu Glu Arg Leu Ile Ser Gln Lys Arg Arg 580 585 590			1776
25	ATG GAG TAA Met Glu			1785

(2) INFORMATION FOR SEQ ID NO:90:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 594 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

35	Met Ile Thr Met Glu Val Leu Cys Met Val Asp Lys Gln Val Ile Glu 1 5 10 15
40	Glu Ile Lys Asn Asn Ala Asn Ile Val Glu Val Ile Gly Asp Val Ile 20 25 30
45	Ser Leu Gln Lys Ala Gly Arg Asn Tyr Leu Gly Leu Cys Pro Phe His 35 40 45
50	Gly Glu Lys Thr Pro Ser Phe Ser Val Val Glu Asp Lys Gln Phe Tyr 50 55 60
55	His Cys Phe Gly Cys Gly Arg Ser Gly Asp Val Phe Lys Phe Ile Glu 65 70 75 80
60	Glu Tyr Gln Gly Val Thr Phe Met Glu Ala Val Gln Ile Leu Gly Gln 85 90 95
65	Arg Val Gly Ile Glu Val Glu Lys Pro Leu Tyr Ser Glu Gln Lys Pro 100 105 110
70	Ala Ser Pro His Gln Ala Leu Tyr Asp Met His Glu Asp Ala Ala Lys 115 120 125
75	Phe Tyr His Ala Ile Leu Met Thr Thr Met Gly Glu Glu Ala Arg 130 135 140

WO 98/26072

PCT/US97/22578

-159-

	Asn	Tyr	Leu	Tyr	Gln	Arg	Gly	Leu	Thr	Asp	Glu	Val	Leu	Lys	His	Phe
	145					150				155						160
5	Trp	Ile	Gly	Leu	Ala	Pro	Pro	Glu	Arg	Asn	Tyr	Leu	Tyr	Gln	Arg	Leu
					165					170					175	
	Ser	Asp	Gln	Tyr	Arg	Glu	Glu	Asp	Leu	Leu	Asp	Ser	Gly	Leu	Phe	Tyr
				180					185					190		
10	Leu	Ser	Asp	Ala	Asn	Gln	Phe	Val	Asp	Thr	Phe	His	Asn	Arg	Ile	Met
			195					200					205			
	Phe	Pro	Leu	Thr	Asn	Asp	Gln	Gly	Lys	Val	Ile	Ala	Phe	Ser	Gly	Arg
		210					215					220				
15	Ile	Trp	Gln	Lys	Thr	Asp	Ser	Gln	Thr	Ser	Lys	Tyr	Lys	Asn	Ser	Arg
		225				230					235				240	
20	Ser	Thr	Val	Ile	Phe	Asn	Lys	Ser	Tyr	Glu	Leu	Tyr	His	Met	Asp	Arg
					245					250					255	
	Ala	Lys	Arg	Ser	Ser	Gly	Lys	Ala	Ser	Glu	Ile	Tyr	Leu	Met	Glu	Gly
				260					265					270		
25	Phe	Met	Asp	Val	Ile	Ala	Ala	Tyr	Arg	Ala	Gly	Ile	Glu	Asn	Ala	Val
			275					280					285			
	Ala	Ser	Met	Gly	Thr	Ala	Leu	Ser	Arg	Glu	His	Val	Glu	His	Leu	Lys
		290					295					300				
30	Arg	Leu	Thr	Lys	Lys	Leu	Val	Leu	Val	Tyr	Asp	Gly	Asp	Lys	Ala	Gly
		305				310					315				320	
35	Gln	Ala	Ala	Thr	Leu	Lys	Ala	Leu	Asp	Glu	Ile	Gly	Asp	Met	Pro	Val
					325					330					335	
	Gln	Ile	Val	Ser	Met	Pro	Asp	Asn	Leu	Asp	Pro	Asp	Glu	Tyr	Leu	Gln
				340					345					350		
40	Lys	Asn	Gly	Pro	Glu	Asp	Leu	Ala	Tyr	Leu	Leu	Thr	Lys	Thr	Arg	Ile
			355				360					365				
	Ser	Pro	Ile	Glu	Phe	Tyr	Ile	His	Gln	Tyr	Lys	Pro	Glu	Asn	Gly	Glu
				370			375					380				
45	Asn	Leu	Gln	Ala	Gln	Ile	Glu	Phe	Leu	Glu	Lys	Ile	Ala	Pro	Leu	Ile
		385				390					395				400	
50	Val	Gln	Glu	Lys	Ser	Ile	Ala	Ala	Gln	Asn	Ser	Tyr	Ile	His	Ile	Leu
				405					410					415		
	Ala	Asp	Ser	Leu	Ala	Ser	Phe	Asp	Tyr	Thr	Gln	Ile	Glu	Gln	Ile	Val
				420					425					430		
55	Asn	Glu	Ser	Arg	Gln	Val	Gln	Arg	Gln	Asn	Arg	Met	Glu	Arg	Ile	Ser
				435					440					445		
	Arg	Pro	Thr	Pro	Ile	Thr	Met	Pro	Val	Thr	Lys	Gln	Leu	Ser	Ala	Ile
		450					455					460				
60	Met	Arg	Ala	Glu	Ala	His	Leu	Leu	Tyr	Arg	Met	Met	Glu	Ser	Pro	Leu

WO 98/26072

PCT/US97/22578

-160-

```

465              470              475              480
Val Leu Asn Asp Tyr Arg Leu Arg Glu Asp Phe Ala Phe Ala Thr Pro
5              485              490              495
Glu Phe Gln Val Leu His Asp Leu Leu Gly Gln Tyr Gly Asn Leu Pro
500              505              510
Pro Glu Val Leu Ala Glu Gln Thr Glu Glu Val Glu Arg Ala Trp Tyr
10              515              520              525
Gln Val Leu Ala Gln Asp Leu Pro Ala Glu Ile Ser Pro Gln Glu Leu
530              535              540
Ser Glu Val Glu Met Thr Arg Asn Lys Ala Leu Leu Asn Gln Asp Asn
15              545              550              555              560
Met Arg Ile Lys Lys Lys Val Gln Glu Ala Ser His Val Gly Asp Thr
20              565              570              575
Asp Thr Ala Leu Glu Glu Leu Glu Arg Leu Ile Ser Gln Lys Arg Arg
580              585              590
Met Glu
25

```

(2) INFORMATION FOR SEQ ID NO:91:

```

30      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 900 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear
35      (ii) MOLECULE TYPE: DNA (genomic)
          (iii) HYPOTHETICAL: NO
          (iv) ANTI-SENSE: NO
40      (ix) FEATURE:
          (A) NAME/KEY: CDS
          (B) LOCATION: 1..897
45      (D) OTHER INFORMATION: Era

```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

```

50  ATG ACT TTT AAA TCA GGC TTT GTA GCC ATT TTA GGA CGT CCC AAT GTT      48
    Met Thr Phe Lys Ser Gly Phe Val Ala Ile Leu Gly Arg Pro Asn Val
      1              5              10              15
55  GGG AAG TCA ACC TTT TTA AAT CAC GTT ATG GGG CAA AAG ATT GCC ATC      96
    Gly Lys Ser Thr Phe Leu Asn His Val Met Gly Gln Lys Ile Ala Ile
      20              25              30
    ATG AGT GAC AAG GCG CAG ACA ACG CGC AAT AAA ATC ATG GGA ATT TAC      144
    Met Ser Asp Lys Ala Gln Thr Thr Arg Asn Lys Ile Met Gly Ile Tyr
      35              40              45

```

PCT/A/S97/22578

	ACG Thr 50	ACT Thr 50	GAT Asp 50	AAG Lys 50	GAG Glu 70	CAA Gln 55	ATT Ile 55	GTC Val 55	TTT Phe 55	ATC Ile 55	GAC Asp 60	ACA Thr 60	CCA Pro 60	GGG Gly 60	ATT Ile 60	CAC His 60	192
5	AAA Lys 65	CCT Pro 65	AAA Lys 65	ACA Thr 65	GCT Ala 70	CTC Leu 70	GGA Gly 70	GAT Asp 70	TTC Phe 70	ATG Met 75	GTT Val 75	GAG Val 75	TCT Ser 75	GCC Ala 75	TAC Tyr 75	AGT Ser 80	240
10	ACC Thr 85	CTT Arg 85	CGC Arg 85	GAA Glu 85	GTG Val 85	GAC Val 85	ACT Thr 85	GTT Val 85	CTT Val 85	TTC Met 90	ATG Met 90	GAG Met 90	CCT Met 90	GCT Met 95	GAT Met 95	GAA Ser 95	288
15	CGC Ala 100	CGT Arg 100	GGT Gly 100	AAG Lys 100	GGG Gly 100	GAC Asp 100	GAT Asp 100	ATT Met 105	ATC Ile 105	ATC Ile 105	GAG Glu 105	CGT Arg 105	CTC Leu 110	AAG Lys 110	GCT Ala 110	GCC Ala 110	336
20	AAG Lys 115	GTT Val 115	CCT Pro 115	GTG Val 115	ATT Ile 115	TTG Leu 120	GTG Val 120	AAT Asn 120	AAA Lys 120	ATC Ile 120	GAT Asp 125	AAG Val 125	GTC Val 125	CAT Val 125	CCA Pro 125	135	384
25	GAC Asp 130	CAG Gln 130	CTC Leu 130	TTG Leu 130	TCT Ser 130	CAG Gln 135	ATT Ile 135	GAT Asp 135	GAC Asp 135	TTC Phe 140	CGT Arg 140	AAT Gln 140	CAA Met 140	ATG Met 140	GAC Asp 140	TTT Phe 140	432
30	AAG Lys 145	GAA Glu 145	ATT Ile 145	GTT Val 145	CCA Pro 150	ATC Pro 150	TCA Ser 150	GCC Ala 150	CTT Leu 150	CAG Gln 155	GGA Asn 155	AAT Asn 155	AAC Val 155	GTG Val 155	TCT Ser 160	CGT Arg 160	480
35	CTA Leu 165	GTG Val 165	GAT Asp 165	ATT Ile 165	TTG Leu 165	AGT Ser 165	GAA Glu 165	AAT Asn 165	CTG Leu 170	GAT Asp 170	GAA Glu 170	GGT Phe 170	TTC Gln 170	CAA Gln 175	TAT Tyr 175	TTT Phe 175	528
40	CCG Met 180	TCT Val 180	GAT Arg 180	CAA Glu 180	ATC Ile 180	ACA Thr 180	GAT Asp 185	CAT His 185	CCA Pro 185	GAA Glu 185	CGT Arg 185	GAT Phe 190	TTC Leu 190	TTA Val 190	GTT Ser 190	TCA Glu 190	576
45	ATG Met 200	GTT Val 200	CGC Arg 195	GAG Glu 195	AAA Lys 200	GTC Val 200	TTG His 200	CAC Leu 200	CTA Thr 200	ACT Thr 200	CGT Arg 200	GAA Glu 205	GAG Glu 205	ATT Ile 205	CCG Pro 205	CAT His 205	624
50	TCT Ser 210	GTA Val 210	GCA Ala 210	GTA Val 210	GTT Val 210	GTT Val 210	GAC Asp 215	TCT Ser 215	ATG Met 215	AAA Lys 215	CGA Arg 220	GAC Asp 220	GAA Glu 220	GAG Glu 220	ACA Thr 220	GAC Asp 220	672
55	AAG Lys 225	GTT Val 225	CAC His 225	ATC Ile 225	CGT Arg 230	GCA Thr 230	ACC Thr 230	ATC Ile 230	ATG Met 230	GTC Val 235	GAG Arg 235	CGC Asp 235	GAT Ser 235	AGC Ser 235	CAA Gln 240	AAA Lys 240	720
60	GGG Gly 250	ATT Ile 250	ATC Ile 250	ATC Ile 250	GGT Gly 245	AAA Lys 245	GGT Gly 245	GGC Gly 245	GCT Ala 250	ATG Met 250	CTT Met 250	AAG Lys 250	AAA Lys 250	ATC Ile 255	GGT Gly 255	AGT Ser 255	768
65	ATG Met 260	GCC Ala 260	CGT Arg 260	CGT Arg 260	GAT Asp 260	ATC Ile 260	GAA Glu 265	CTC Leu 265	ATG Met 265	CTA Val 265	GGA Arg 265	GAC Asp 270	AAG Val 270	GTC Val 270	TTC Phe 270	CTA Leu 270	816
70	GAA Glu 275	ACC Thr 275	TGG Val 275	GTC Lys 275	CTC Lys 275	AAG Lys 275	AAA Lys 280	AAC Trp 280	TGG Trp 280	CGC Arg 280	GAT Asp 285	AAA Lys 285	AAG Lys 285	CTA Lys 285	GAT Leu 285	ASP Asp 285	864
75	TTG Thr 290	GCT Thr 290	GAC Thr 290	TTT Thr 290	GGC Thr 290	TAT Thr 290	AAT Thr 290	AGA Thr 290	GAA Thr 290	TAC Thr 290	TAA Thr 290						900

WO 98/26072

PCT/US97/22578

-162-

Leu Ala Asp Phe Gly Tyr Asn Glu Arg Glu Tyr  
290 295

5 (2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 299 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

15

Met Thr Phe Lys Ser Gly Phe Val Ala Ile Leu Gly Arg Pro Asn Val  
1 5 10 15

20

Gly Lys Ser Thr Phe Leu Asn His Val Met Gly Gln Lys Ile Ala Ile  
20 25 30

Met Ser Asp Lys Ala Gln Thr Thr Arg Asn Lys Ile Met Gly Ile Tyr  
35 40 45

25

Thr Thr Asp Lys Glu Gln Ile Val Phe Ile Asp Thr Pro Gly Ile His  
50 55 60

30

Lys Pro Lys Thr Ala Leu Gly Asp Phe Met Val Glu Ser Ala Tyr Ser  
65 70 75 80

Thr Leu Arg Glu Val Asp Thr Val Leu Phe Met Val Pro Ala Asp Glu  
85 90 95

35

Ala Arg Gly Lys Gly Asp Asp Met Ile Ile Glu Arg Leu Lys Ala Ala  
100 105 110

Lys Val Pro Val Ile Leu Val Val Asn Lys Ile Asp Lys Val His Pro  
115 120 125

40

Asp Gln Leu Leu Ser Gln Ile Asp Asp Phe Arg Asn Gln Met Asp Phe  
130 135 140

Lys Glu Ile Val Pro Ile Ser Ala Leu Gln Gly Asn Asn Val Ser Arg  
145 150 155 160

45

Leu Val Asp Ile Leu Ser Glu Asn Leu Asp Glu Gly Phe Gln Tyr Phe  
165 170 175

50

Pro Ser Asp Gln Ile Thr Asp His Pro Glu Arg Phe Leu Val Ser Glu  
180 185 190

Met Val Arg Glu Lys Val Leu His Leu Thr Arg Glu Glu Ile Pro His  
195 200 205

55

Ser Val Ala Val Val Val Asp Ser Met Lys Arg Asp Glu Glu Thr Asp  
210 215 220

Lys Val His Ile Arg Ala Thr Ile Met Val Glu Arg Asp Ser Gln Lys  
225 230 235 240

60

Gly Ile Ile Ile Gly Lys Gly Gly Ala Met Leu Lys Lys Ile Gly Ser



WO 98/26072

PCT/US97/22578

-163-

	245	250	255	
	Met Ala Arg Arg Asp Ile Glu Leu Met Leu Gly Asp Lys Val Phe Leu			
	260	265	270	
5	Glu Thr Trp Val Lys Val Lys Lys Asn Trp Arg Asp Lys Lys Leu Asp			
	275	280	285	
10	Leu Ala Asp Phe Gly Tyr Asn Glu Arg Glu Tyr			
	290	295		
(2) INFORMATION FOR SEQ ID NO:93:				
15	(i) SEQUENCE CHARACTERISTICS:			
	(A) LENGTH: 1011 base pairs			
	(B) TYPE: nucleic acid			
20	(C) STRANDEDNESS: single			
	(D) TOPOLOGY: linear			
	(ii) MOLECULE TYPE: DNA (genomic)			
	(iii) HYPOTHETICAL: NO			
25	(iv) ANTI-SENSE: NO			
	(ix) FEATURE:			
30	(A) NAME/KEY: CDS			
	(B) LOCATION: 1..1008			
	(D) OTHER INFORMATION: Gcp			
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:			
	ATG AAG GAT AGA TAT ATT TTA GCA TTT GAG ACA TCC TGT GAT GAG ACC	48		
	Met Lys Asp Arg Tyr Ile Leu Ala Phe Glu Thr Ser Cys Asp Glu Thr			
	1 5 10 15			
40	AGT GTC GCC GTC TTG AAA AAC GAC GAT GAG CTC TTG TCC AAT GTC ATT	96		
	Ser Val Ala Val Leu Lys Asn Asp Asp Glu Leu Leu Ser Asn Val Ile			
	20 25 30			
45	GCT AGT CAA ATT GAG AGT CAC AAA CGT TTT GGT GGC GTA GTG CCC GAA	144		
	Ala Ser Gln Ile Glu Ser His Lys Arg Phe Gly Gly Val Val Pro Glu			
	35 40 45			
50	GTA GCC AGT CGT CAC CAT GTC GAG GTC ATT ACA GCC TGT ATC GAG GAG	192		
	Val Ala Ser Arg His His Val Glu Val Ile Thr Ala Cys Ile Glu Glu			
	50 55 60			
55	GCA TTG GCA GAA GCA GGG ATT ACC GAA GAG GAC GTG ACA GCT GTT GCG	240		
	Ala Leu Ala Glu Ala Gly Ile Thr Glu Glu Asp Val Thr Ala Val Ala			
	65 70 75 80			
60	GTT ACC TAC GGA CCA GGC TTG GTC GGA GCC TTG CTA GTT GGT TTG TCA	288		
	Val Thr Tyr Gly Pro Gly Leu Val Gly Ala Leu Leu Val Gly Leu Ser			
	85 90 95			
60	GCT GCC AAG GCC TTT GCT TGG GCT CAC GGA CTT CCA CTG ATT CCT GTT	336		
	Ala Ala Lys Ala Phe Ala Trp Ala His Gly Leu Pro Leu Ile Pro Val			



WO 98/26072

PCT/US97/22578

-165-

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 336 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

Met Lys Asp Arg Tyr Ile Leu Ala Phe Glu Thr Ser Cys Asp Glu Thr  
1 5 10 15  
Ser Val Ala Val Leu Lys Asn Asp Asp Glu Leu Leu Ser Asn Val Ile  
20 25 30  
Ala Ser Gln Ile Glu Ser His Lys Arg Phe Gly Gly Val Val Pro Glu  
35 40 45  
Val Ala Ser Arg His His Val Glu Val Ile Thr Ala Cys Ile Glu Glu  
50 55 60  
Ala Leu Ala Glu Ala Gly Ile Thr Glu Glu Asp Val Thr Ala Val Ala  
65 70 75 80  
Val Thr Tyr Gly Pro Gly Leu Val Gly Ala Leu Leu Val Gly Leu Ser  
85 90 95  
Ala Ala Lys Ala Phe Ala Trp Ala His Gly Leu Pro Leu Ile Pro Val  
100 105 110  
Asn His Met Ala Gly His Leu Met Ala Ala Gln Ser Val Glu Pro Leu  
115 120 125  
Glu Phe Pro Leu Leu Ala Leu Leu Val Ser Gly Gly His Thr Glu Leu  
130 135 140  
Val Tyr Val Ser Glu Ala Gly Asp Tyr Lys Ile Val Gly Glu Thr Arg  
145 150 155 160  
Asp Asp Ala Val Gly Glu Ala Tyr Asp Lys Val Gly Arg Val Met Gly  
165 170 175  
Leu Thr Tyr Pro Ala Gly Arg Glu Ile Asp Glu Leu Ala His Gln Gly  
180 185 190  
His Asp Ile Tyr Asp Phe Pro Arg Ala Met Ile Lys Glu Asp Asn Leu  
195 200 205  
Glu Phe Ser Phe Ser Gly Leu Lys Ser Ala Phe Ile Asn Leu His His  
210 215 220  
Asn Ala Glu Gln Lys Gly Glu Ser Leu Ser Thr Glu Asp Leu Cys Ala  
225 230 235 240  
Ser Phe Gln Ala Ala Val Met Asp Ile Leu Met Ala Lys Thr Lys Lys  
245 250 255  
Ala Leu Glu Lys Tyr Pro Val Lys Thr Leu Val Val Ala Gly Gly Val  
260 265 270

WO 98/26072

PCT/US97/22578

-166-

	Ala	Ala	Asn	Lys	Gly	Leu	Arg	Glu	Arg	Leu	Ala	Thr	Glu	Ile	Thr	Asp	
			275					280					285				
5	Val	Asn	Val	Ile	Ile	Pro	Pro	Leu	Arg	Leu	Cys	Gly	Asp	Asn	Ala	Gly	
		290					295				300						
	Met	Ile	Ala	Tyr	Ala	Ser	Val	Ser	Glu	Trp	Asn	Lys	Glu	Asn	Phe	Ala	
		305			310						315					320	
10	Asn	Leu	Asp	Leu	Asn	Ala	Lys	Pro	Ser	Leu	Ala	Phe	Asp	Thr	Met	Glu	
				325						330					335		
	(2) INFORMATION FOR SEQ ID NO:95:																
15	(i) SEQUENCE CHARACTERISTICS:																
	(A) LENGTH: 774 base pairs																
	(B) TYPE: nucleic acid																
20	(C) STRANDEDNESS: single																
	(D) TOPOLOGY: linear																
	(ii) MOLECULE TYPE: DNA (genomic)																
	(iii) HYPOTHETICAL: NO																
25	(iv) ANTI-SENSE: NO																
	(ix) FEATURE:																
30	(A) NAME/KEY: CDS																
	(B) LOCATION: 1..771																
	(D) OTHER INFORMATION: HI0454																
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:																
	ATG	ATT	TTT	GAT	ACA	CAT	ACA	CAC	TTG	AAT	GTA	GAA	GAA	TTT	GCA	GGT	48
	Met	Ile	Phe	Asp	Thr	His	Thr	His	Leu	Asn	Val	Glu	Glu	Phe	Ala	Gly	
40	1				5				10					15			
	CCT	GAG	GCA	GAA	GAA	ATT	GCC	TTG	GCT	GCT	GAG	ATG	GGT	GTG	ACA	CAG	96
	Arg	Glu	Ala	Glu	Glu	Ile	Ala	Leu	Ala	Ala	Glu	Met	Gly	Val	Thr	Gln	
				20				25					30				
45	ATG	AAT	ATT	GTT	GGT	TTT	GAT	AAA	CCG	ACG	ATT	GAG	CAT	GCC	TTG	GAG	144
	Met	Asn	Ile	Val	Gly	Phe	Asp	Lys	Pro	Thr	Ile	Glu	His	Ala	Leu	Glu	
			35				40					45					
50	TTG	GTA	GAT	GAG	TAT	GAG	CAG	CTC	TAT	GCG	ACT	ATT	GGT	TGG	CAT	CCT	192
	Leu	Val	Asp	Glu	Tyr	Glu	Gln	Leu	Tyr	Ala	Thr	Ile	Gly	Trp	His	Pro	
		50					55					60					
55	ACA	GAA	GCT	GGT	ACT	TAT	ACA	GAG	GAA	GTT	GAG	GCT	TAC	TTG	TTG	GAT	240
	Thr	Glu	Ala	Gly	Thr	Tyr	Thr	Glu	Glu	Val	Glu	Ala	Tyr	Leu	Leu	Asp	
		65			70					75				80			
60	AAG	TTA	AAA	CAT	TCC	AAG	GTT	GTG	GCT	TTA	GGT	GAA	ATT	GGC	TTA	GAC	288
	Lys	Leu	Lys	His	Ser	Lys	Val	Val	Ala	Leu	Gly	Glu	Ile	Gly	Leu	Asp	
				85					90				95				
	TAC	CAT	TGG	ATG	ACA	GCG	CCC	AAA	GAG	GTG	CAG	GAG	CAG	GTT	TTT	CGC	336

WO 98/26072

PCT/US97/22578

-167-

	Tyr	His	Trp	Met	Thr	Ala	Pro	Lys	Glu	Val	Gln	Glu	Gln	Val	Phe	Arg	
				100					105					110			
5	CGT	CAG	ATT	CAG	CTA	TCT	AAG	GAC	TTG	GAT	TTG	CCT	TTT	GTT	GTC	CAT	364
	Arg	Gln	Ile	Gln	Leu	Ser	Lys	Asp	Leu	Asp	Leu	Pro	Phe	Val	Val	His	
				115				120					125				
10	ACC	CGT	GAT	GCG	CTG	GAA	GAT	ACC	TAT	GAG	ATT	ATC	AAG	AGT	GAG	GCC	432
	Thr	Arg	Asp	Ala	Leu	Glu	Asp	Thr	Tyr	Glu	Ile	Ile	Lys	Ser	Glu	Gly	
				130			135					140					
15	GTT	GGT	CCT	CGT	GGT	ATC	ATG	CAT	TCA	TTT	TCA	GGG	ACG	CTT	GAG		480
	Val	Gly	Pro	Arg	Gly	Ile	Met	His	Ser	Phe	Ser	Gly	Thr	Leu	Glu		
					145		150			155					160		
20	TGG	GCA	GAG	AAG	TTT	GTG	GAT	CTT	GGT	ATG	ACC	ATT	TCC	TTC	TCA	GGA	528
	Trp	Ala	Glu	Lys	Phe	Val	Asp	Leu	Gly	Met	Thr	Ile	Ser	Phe	Ser	Gly	
					165				170						175		
25	GTG	GTG	ACC	TTC	AAG	AAG	GCA	ACT	GAC	CTC	CAA	GAA	GCA	GCT	AAA	GAG	576
	Val	Val	Thr	Phe	Lys	Lys	Ala	Thr	Asp	Leu	Gln	Glu	Ala	Ala	Lys	Glu	
				180					185					190			
30	TTA	CCT	TTG	GAC	AAG	ATG	TTG	GTA	GAA	ACA	GAT	GCG	CCT	TAC	TTA	GCA	624
	Leu	Pro	Leu	Asp	Lys	Met	Leu	Val	Glu	Thr	Asp	Ala	Pro	Tyr	Leu	Ala	
				195			200					205					
35	CCT	GTA	CCC	AAG	CGT	GGT	CGT	GAA	AAT	AAA	ACA	GCC	TAT	ACT	CGC	TAT	672
	Pro	Val	Pro	Lys	Arg	Gly	Arg	Glu	Asn	Lys	Thr	Ala	Tyr	Thr	Arg	Tyr	
				210			215					220					
40	GTG	GTC	GAC	TTT	ATC	GCT	GAC	TTG	CGT	GGT	ATG	ACG	ACA	GAA	GAG	CTG	720
	Val	Val	Asp	Phe	Ile	Ala	Asp	Leu	Arg	Gly	Met	Thr	Thr	Glu	Glu	Leu	
				225		230				235					240		
45	GCG	GTA	GCA	ACG	ACT	GCA	AAT	GCA	GAA	CGC	ATT	TTT	GGA	TTG	GAC	AGC	768
	Ala	Val	Ala	Thr	Thr	Ala	Asn	Ala	Glu	Arg	Ile	Phe	Gly	Leu	Asp	Ser	
					245				250					255			
50	AAG	TAA															774
	Lys																
55	(2) INFORMATION FOR SEQ ID NO:96:																
	(i) SEQUENCE CHARACTERISTICS:																
	(A) LENGTH: 257 amino acids																
	(B) TYPE: amino acid																
	(C) TOPOLOGY: linear																
	(ii) MOLECULE TYPE: protein																
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:																
60	Met	Ile	Phe	Asp	Thr	His	Thr	His	Leu	Asn	Val	Glu	Glu	Phe	Ala	Gly	
	1				5					10					15		
	Arg	Glu	Ala	Glu	Glu	Ile	Ala	Leu	Ala	Ala	Glu	Met	Gly	Val	Thr	Gln	
				20				25						30			

WO 98/26072

PCT/US97/22578

-168-

```

Met Asn Ile Val Gly Phe Asp Lys Pro Thr Ile Glu His Ala Leu Glu
      35              40              45
5  Leu Val Asp Glu Tyr Glu Gln Leu Tyr Ala Thr Ile Gly Trp His Pro
      50              55              60
    Thr Glu Ala Gly Thr Tyr Thr Glu Glu Val Glu Ala Tyr Leu Leu Asp
      65              70              75              80
10 Lys Leu Lys His Ser Lys Val Val Ala Leu Gly Glu Ile Gly Leu Asp
      85              90              95
    Tyr His Trp Met Thr Ala Pro Lys Glu Val Gln Gln Val Phe Arg
      100             105             110
15 Arg Gln Ile Gln Leu Ser Lys Asp Leu Asp Leu Pro Phe Val Val His
      115             120             125
    Thr Arg Asp Ala Leu Glu Asp Thr Tyr Glu Ile Ile Lys Ser Glu Gly
      130             135             140
20 Val Gly Pro Arg Gly Gly Ile Met His Ser Phe Ser Gly Thr Leu Glu
      145             150             155             160
25 Trp Ala Glu Lys Phe Val Asp Leu Gly Met Thr Ile Ser Phe Ser Gly
      165             170             175
    Val Val Thr Phe Lys Lys Ala Thr Asp Leu Gln Glu Ala Ala Lys Glu
      180             185             190
30 Leu Pro Leu Asp Lys Met Leu Val Glu Thr Asp Ala Pro Tyr Leu Ala
      195             200             205
    Pro Val Pro Lys Arg Gly Arg Glu Asn Lys Thr Ala Tyr Thr Arg Tyr
      210             215             220
35 Val Val Asp Phe Ile Ala Asp Leu Arg Gly Met Thr Thr Glu Glu Leu
      225             230             235             240
40 Ala Val Ala Thr Thr Ala Asn Ala Glu Arg Ile Phe Gly Leu Asp Ser
      245             250             255
    Lys

```

45

(2) INFORMATION FOR SEQ ID NO:97:

50

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1959 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

55

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

60

(ix) FEATURE:

PCT/US97/22578

(A) NAME/KEY: CDS  
(B) LOCATION: 1..1959  
(D) OTHER INFORMATION: Ligase

5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:																			
	ATG	AAT	AAA	AGA	ATG	AAT	GAG	TTA	GTC	GCT	TTG	CTC	AAT	CGC	TAT	GCG	48			
10	Met	Asn	Lys	Arg	Met	Asn	Glu	Leu	Val	Ala	Leu	Leu	Asn	Arg	Tyr	Ala				
	1				5					10					15					
	ACT	GAG	TAC	TAT	ACC	AGC	GAT	AAT	CCC	TCG	GTT	TCA	GAC	AGT	GAG	TAT	96			
	Thr	Glu	Tyr	Tyr	Thr	Ser	Asp	Asn	Pro	Ser	Val	Ser	Asp	Ser	Glu	Tyr				
				20					25					30						
15	GAC	CGC	CTT	TAC	CGT	GAG	TTG	GTC	GAG	TTA	GAA	ACT	GCT	TAT	CCA	GAG	144			
	Asp	Arg	Leu	Tyr	Arg	Glu	Leu	Val	Glu	Leu	Glu	Thr	Ala	Tyr	Pro	Glu				
			35					40					45							
20	CAA	GTG	CTA	GCA	GAC	AGT	CCG	ACT	CAT	CGT	GTT	GGT	GGC	AAG	GTT	TTA	192			
	Gln	Val	Leu	Ala	Asp	Ser	Pro	His	Arg	Thr	Gly	Gly	Ala	Lys	Val	Leu				
		50					55					60								
	GAT	GGT	TTT	GAA	AAA	TAC	AGT	CAT	CAG	TAT	CCT	CTT	TAT	AGT	TTG	CAG	240			
25	Asp	Gly	Phe	Glu	Lys	Tyr	Ser	His	Gln	Tyr	Pro	Gly	Thr	Tyr	Ser	Leu				
		65				70					75					80				
	GAT	GCT	TTT	TCA	CGT	GAG	GAG	CTA	GAT	GCT	TTT	GAT	GCG	CGT	GTT	CGT	288			
30	Asp	Ala	Phe	Ser	Arg	Glu	Glu	Leu	Asp	Ala	Phe	Asp	Ala	Arg	Val	Arg				
					85					90					95					
	AAG	GAA	GTG	GCT	CAT	CCG	ACC	TAT	ATT	TGT	GAG	CTG	AAA	ATC	GAT	GGC	336			
	Lys	Glu	Val	Ala	Ala	His	Pro	Thr	Tyr	Ile	Cys	Glu	Leu	Lys	Ile	Asp				
				100						105				110						
35	TTA	TCT	ATC	TCG	CTG	ACT	TAT	GAA	AAG	GGG	ATT	TTG	GTT	GCT	GGG	GTA	384			
	Leu	Ser	Ile	Ser	Leu	Thr	Tyr	Glu	Lys	Gly	Ile	Leu	Val	Ala	Gly	Val				
			115					120					125							
40	ACA	CGT	GGA	GAT	GGT	TCA	ATT	GGT	GAA	AAT	ATC	ACA	GAA	AAC	CTC	AAG	432			
	Thr	Arg	Gly	Asp	Gly	Ser	Ile	Gly	Glu	Asn	Ile	Thr	Glu	Asn	Leu	Lys				
		130				135						140								
	CGT	GTT	AAG	GAC	ATC	CCT	TTG	ACT	TTG	CCA	GAA	GAA	CTA	GAT	ATC	ACA	480			
45	Arg	Val	Lys	Val	Asp	Ile	Pro	Leu	Thr	Pro	Glu	Glu	Leu	Asp	Ile	Thr				
		145				150					155				160					
	GTT	CGT	GGG	GAA	TGT	TAC	ATG	CCA	CGC	GCT	TCC	TTT	GAC	CAA	GTT	AAC	528			
50	Val	Arg	Gly	Glu	Cys	Tyr	Met	Pro	Arg	Ala	Ser	Phe	Asp	Gln	Val	Asn				
				165						170				175						
	CAA	GCG	CGC	CAA	GAA	AAT	GGA	GAG	CCT	GAA	TTT	GCT	AAT	CCT	CGT	AAT	576			
	Gln	Ala	Arg	Gln	Glu	Asn	Gly	Glu	Pro	Glu	Phe	Ala	Asn	Pro	Arg	Asn				
				180					185					190						





PCT/US97/22578

5	GAA GAG GAT TTC CTC CTT TTA GAG GGC GTT AAG GAA AAG TCC GCT GCT Glu Glu Asp Phe Leu Leu Leu Leu Leu Gly Val Lys Lys Glu Lys Ser Ala Ala 465 470 475	1440
	AAA CTG TAT CAG GCT ATC CAA GCA TCA AAG GAA AAT TCT GCC GAG AAG Lys Leu Tyr Gln Ala Ile Gln Ala Ser Lys Lys Glu Asn Ser Ala Glu Lys 485 490 495	1488
10	CTC TTA TTT GGT TTG GGA ATT CGT CAT ATT GTC GGA AGC AAG GCT AGT CAG Leu Leu Phe Gly Leu Gly Ile Arg His Val Lys Arg Lys Ala Ser Gln 500 505	1536
15	CTT TTA CTT CAA TAT TTC CAT TCA ATT GAA AAT CTG TAT CAG GCA GAT Leu Leu Leu Leu Gln Tyr Phe His Ser Ile Glu Asn Leu Tyr 515 520 525	1584
20	TCA GAG GAA GTG GCT AGT ATT GAA AGT CTA GGT GGC GTG ATT GCC AAA Ser Glu Glu Val Ala Ser Ile Glu Ser Leu Gly Gly Val Ile Ala Lys 530 535 540	1632
25	AGT CTT CAG ACT TAT TTT GCG GCA GAA GGC TCT GAA ATT CTG CTC AGA Ser Leu Gln Thr Tyr Phe Ala Ala Glu Gly Ser Glu Ile Leu Leu Arg 545 550 555	1680
30	GAA TTG AAA GAA ACT GGG GTC AAT CTG GAC TAT AAA GGA CAG ACG GTA Glu Leu Lys Lys Glu Thr Gly Val Asn Leu Asp Tyr Lys Gly Gln Thr Val 565 570 575	1728
35	GTA GCG GAT GCG GCC TTG TCA GGT TTG ACC GTG GTA TTG ACA GGA AAA Val Ala Asp Ala Ala Ser Gly Leu Thr Val Val Leu Thr Gly Lys 580 585 590	1776
40	TTG GAA CGA CTC AAG GCG TCA GAA GCT AAA AGT AAA CTC GAA AGT CTG Leu Glu Arg Leu Lys Arg Ser Glu Ala Lys Ser Lys Leu Glu Ser Leu 595 600 605	1824
45	GGT GCC AAA GTG ACA GGT AGT GTT TCT AAA AAG ACC GAC CTC GTC GTG Gly Ala Lys Val Thr Gly Ser Val Ser Lys Lys Thr Asp Leu Val Val 610 615 620	1872
50	GTA GGT GCA GAC GCT GGA AGT AAA CTG CAA AAA GCA CAA GAA CTT GGT Val Gly Ala Asp Ala Gly Ser Lys Leu Gln Lys Ala Gln Glu Leu Gly 625 630 635 640	1920
55	ATC CAG GTC AGA CAT GAG GCA TGG CTA GAA ACT TTG TAA Ile Gln Val Arg Asp Glu Ala Trp Leu Glu Ser Leu 645 650	1959
60	(2) INFORMATION FOR SEQ ID NO:98:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 653 amino acids	
	(B) TYPE: amino acid	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:	

WO 98/26072

PCT/US97/22578

-172-

	Met	Asn	Lys	Arg	Met	Asn	Glu	Leu	Val	Ala	Leu	Leu	Asn	Arg	Tyr	Ala
	1				5					10					15	
5	Thr	Glu	Tyr	Tyr	Thr	Ser	Asp	Asn	Pro	Ser	Val	Ser	Asp	Ser	Glu	Tyr
				20					25					30		
	Asp	Arg	Leu	Tyr	Arg	Glu	Leu	Val	Glu	Leu	Glu	Thr	Ala	Tyr	Pro	Glu
				35				40					45			
10	Gln	Val	Leu	Ala	Asp	Ser	Pro	Thr	His	Arg	Val	Gly	Gly	Lys	Val	Leu
		50					55					60				
	Asp	Gly	Phe	Glu	Lys	Tyr	Ser	His	Gln	Tyr	Pro	Leu	Tyr	Ser	Leu	Gln
	65					70					75				80	
15	Asp	Ala	Phe	Ser	Arg	Glu	Glu	Leu	Asp	Ala	Phe	Asp	Ala	Arg	Val	Arg
					85					90				95		
	Lys	Glu	Val	Ala	His	Pro	Thr	Tyr	Ile	Cys	Glu	Leu	Lys	Ile	Asp	Gly
				100					105					110		
20	Leu	Ser	Ile	Ser	Leu	Thr	Tyr	Glu	Lys	Gly	Ile	Leu	Val	Ala	Gly	Val
			115					120					125			
25	Thr	Arg	Gly	Asp	Gly	Ser	Ile	Gly	Glu	Asn	Ile	Thr	Glu	Asn	Leu	Lys
		130					135					140				
	Arg	Val	Lys	Asp	Ile	Pro	Leu	Thr	Leu	Pro	Glu	Glu	Leu	Asp	Ile	Thr
		145				150					155				160	
30	Val	Arg	Gly	Glu	Cys	Tyr	Met	Pro	Arg	Ala	Ser	Phe	Asp	Gln	Val	Asn
					165				170					175		
	Gln	Ala	Arg	Gln	Glu	Asn	Gly	Glu	Pro	Glu	Phe	Ala	Asn	Pro	Arg	Asn
				180					185					190		
35	Ala	Ala	Ala	Gly	Thr	Leu	Arg	Gln	Leu	Asp	Thr	Ala	Val	Val	Ala	Lys
			195					200					205			
40	Arg	Asn	Leu	Ala	Thr	Phe	Leu	Tyr	Gln	Glu	Ala	Ser	Pro	Ser	Thr	Arg
		210					215					220				
	Asp	Ser	Gln	Glu	Lys	Gly	Leu	Lys	Tyr	Leu	Glu	Gln	Leu	Gly	Phe	Val
		225				230				235				240		
45	Val	Asn	Pro	Lys	Arg	Ile	Leu	Ala	Glu	Asn	Ile	Asp	Glu	Ile	Trp	Asn
				245					250					255		
	Phe	Ile	Gln	Glu	Val	Gly	Gln	Glu	Arg	Glu	Asn	Leu	Pro	Tyr	Asp	Ile
			260					265					270			
50	Asp	Gly	Val	Val	Ile	Lys	Val	Asn	Asp	Leu	Ala	Ser	Gln	Glu	Glu	Leu
			275					280					285			
	Gly	Phe	Thr	Val	Lys	Ala	Pro	Lys	Trp	Ala	Val	Ala	Tyr	Lys	Phe	Pro
			290				295					300				
	Ala	Glu	Glu	Lys	Glu	Ala	Gln	Leu	Leu	Ser	Val	Asp	Trp	Thr	Val	Gly
		305				310					315				320	
60	Arg	Thr	Gly	Val	Val	Thr	Pro	Thr	Ala	Asn	Leu	Thr	Pro	Val	Gln	Leu

WO 98/26072

PCT/US97/22578

-173-

		325		330		335
	Ala Gly Thr Thr Val Ser Arg Ala Thr Leu His Asn Val Asp Tyr Ile					
		340		345		350
5	Ala Glu Lys Asp Ile Arg Lys Asp Asp Thr Val Ile Val Tyr Lys Ala					
		355		360		365
10	Gly Asp Ile Ile Pro Ala Val Leu Arg Val Val Glu Ser Lys Arg Val					
		370		375		380
	Ser Glu Glu Lys Leu Asp Ile Pro Thr Asn Cys Pro Ser Cys Asn Ser					
		385		390		400
15	Asp Leu Leu His Phe Glu Asp Glu Val Ala Leu Arg Cys Ile Asn Pro					
		405		410		415
	Arg Cys Pro Ala Gln Ile Met Glu Gly Leu Ile His Phe Ala Ser Arg					
		420		425		430
20	Asp Ala Met Asn Ile Thr Gly Leu Gly Pro Ser Ile Val Glu Lys Leu					
		435		440		445
25	Phe Ala Ala Asn Leu Val Lys Asp Val Ala Asp Ile Tyr Arg Leu Gln					
		450		455		460
	Glu Glu Asp Phe Leu Leu Leu Glu Gly Val Lys Glu Lys Ser Ala Ala					
		465		470		475
30	Lys Leu Tyr Gln Ala Ile Gln Ala Ser Lys Glu Asn Ser Ala Glu Lys					
		485		490		495
	Leu Leu Phe Gly Leu Gly Ile Arg His Val Gly Ser Lys Ala Ser Gln					
		500		505		510
35	Leu Leu Leu Gln Tyr Phe His Ser Ile Glu Asn Leu Tyr Gln Ala Asp					
		515		520		525
40	Ser Glu Glu Val Ala Ser Ile Glu Ser Leu Gly Gly Val Ile Ala Lys					
		530		535		540
	Ser Leu Gln Thr Tyr Phe Ala Ala Glu Gly Ser Glu Ile Leu Leu Arg					
		545		550		555
45	Glu Leu Lys Glu Thr Gly Val Asn Leu Asp Tyr Lys Gly Gln Thr Val					
		565		570		575
	Val Ala Asp Ala Ala Leu Ser Gly Leu Thr Val Val Leu Thr Gly Lys					
		580		585		590
50	Leu Glu Arg Leu Lys Arg Ser Glu Ala Lys Ser Lys Leu Glu Ser Leu					
		595		600		605
55	Gly Ala Lys Val Thr Gly Ser Val Ser Lys Lys Thr Asp Leu Val Val					
		610		615		620
	Val Gly Ala Asp Ala Gly Ser Lys Leu Gln Lys Ala Gln Glu Leu Gly					
		625		630		635
60	Ile Gln Val Arg Asp Glu Ala Trp Leu Glu Ser Leu					
		645		650		

WO 98/26072

PCT/US97/22578

-174-

(2) INFORMATION FOR SEQ ID NO:99:

5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 981 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

15 (iv) ANTI-SENSE: NO

(ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 1..981  
 (D) OTHER INFORMATION: Mray

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

25	ATG TTT ATT TCC ATC AGT GCT GGA ATT GTG ACA TTT TTA CTA ACT TTA	48
	Met Phe Ile Ser Ile Ser Ala Gly Ile Val Thr Phe Leu Leu Thr Leu	
	1 5 10 15	
30	GTA GGA ATT CCG GCC TTT ATC CAA TTT TAT AGA AAG GCG CAA ATT ACA	96
	Val Gly Ile Pro Ala Phe Ile Gln Phe Tyr Arg Lys Ala Gln Ile Thr	
	20 25 30	
35	GGC CAG CAG ATG CAT CAG GAT GTC AAA CAG CAT CAG GCA AAA GCT GGG	144
	Gly Gln Gln Met His Glu Asp Val Lys Gln His Gln Ala Lys Ala Gly	
	35 40 45	
40	ACT CCT ACA ATG GGA GGT TTG GTT TTC TTG ATT ACT TCT GTT TTG GTT	192
	Thr Pro Thr Met Gly Gly Leu Val Phe Leu Ile Thr Ser Val Leu Val	
	50 55 60	
45	GCT TTC TTT TTC GCC CTA TTT AGT AGC CAA TTC AGC AAT AAT GTG GGA	240
	Ala Phe Phe Phe Ala Leu Phe Ser Ser Gln Phe Ser Asn Asn Val Gly	
	65 70 75 80	
50	ATG ATT TTG TTC ATC TTG GTC TTG TAT GGC TTG GTC GGA TTT TTA GAT	288
	Met Ile Leu Phe Ile Leu Val Leu Tyr Gly Gly Leu Val Gly Phe Leu Asp	
	85 90 95	
55	GAC TTT CTC AAG GTC TTT CGT AAA ATC AAT GAG GGG CTT AAT CCT AAG	336
	Asp Phe Leu Lys Val Phe Arg Lys Ile Asn Glu Gly Leu Asn Pro Lys	
	100 105 110	
60	CAA AAA TTA GCT CTT CAG CTT CTA GGT GGA GTT ATC TTC TAT CTT TTC	384
	Gln Lys Leu Ala Leu Gln Leu Leu Gly Gly Val Ile Phe Tyr Leu Phe	
	115 120 125	
65	TAT GAG CGC GGT GGC GAT ATC CTG TCT GTC TTT GGT TAT CCA GTT CAT	432
	Tyr Glu Arg Gly Gly Asp Ile Leu Ser Val Phe Gly Tyr Pro Val His	
	130 135 140	
70	TTG GGA TTT TTC TAT ATT TTC TTC GCT CTT TTC TGG CTA GTC GGT TTT	480
	Leu Gly Phe Phe Tyr Ile Phe Phe Ala Leu Phe Trp Leu Val Gly Phe	

WO 98/26072

PCT/US97/22578

-175-

[illegible]

50 (2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 326 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

55 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

60 Met Phe Ile Ser Ile Ser Ala Gly Ile Val Thr Phe Leu Leu Thr Leu  
1 5 10 15

WO 98/26072

PCT/US97/22578

-176-

Val Gly Ile Pro Ala Phe Ile Gln Phe Tyr Arg Lys Ala Gln Ile Thr  
20 25 30

5 Gly Gln Gln Met His Glu Asp Val Lys Gln His Gln Ala Lys Ala Gly  
35 40 45

Thr Pro Thr Met Gly Gly Leu Val Phe Leu Ile Thr Ser Val Leu Val  
50 55 60

10 Ala Phe Phe Phe Ala Leu Phe Ser Ser Gln Phe Ser Asn Asn Val Gly  
65 70 75 80

15 Met Ile Leu Phe Ile Leu Val Leu Tyr Gly Leu Val Gly Phe Leu Asp  
85 90 95

Asp Phe Leu Lys Val Phe Arg Lys Ile Asn Glu Gly Leu Asn Pro Lys  
100 105 110

20 Gln Lys Leu Ala Leu Gln Leu Leu Gly Gly Val Ile Phe Tyr Leu Phe  
115 120 125

Tyr Glu Arg Gly Gly Asp Ile Leu Ser Val Phe Gly Tyr Pro Val His  
130 135 140

25 Leu Gly Phe Phe Tyr Ile Phe Phe Ala Leu Phe Trp Leu Val Gly Phe  
145 150 155 160

30 Ser Asn Ala Val Asn Leu Thr Asp Gly Val Asp Gly Leu Ala Ser Ile  
165 170 175

Ser Val Val Ile Ser Leu Phe Ala Tyr Gly Val Ile Ala Tyr Val Gln  
180 185 190

35 Gly Gln Met Asp Ile Leu Leu Val Ile Leu Ala Met Ile Gly Gly Leu  
195 200 205

Leu Gly Phe Phe Ile Phe Asn His Lys Pro Ala Lys Val Phe Met Gly  
210 215 220

40 Asp Val Gly Ser Leu Ala Leu Gly Gly Met Leu Ala Ala Ile Ser Met  
225 230 235 240

45 Ala Leu His Gln Glu Trp Thr Leu Leu Ile Ile Gly Ile Val Tyr Val  
245 250 255

Phe Glu Thr Thr Ser Val Met Met Gln Val Ser Tyr Phe Lys Leu Thr  
260 265 270

50 Gly Gly Lys Arg Ile Phe Arg Met Thr Pro Val His His His Phe Glu  
275 280 285

Leu Gly Gly Leu Ser Gly Lys Gly Asn Pro Trp Ser Glu Trp Lys Val  
290 295 300

55 Asp Phe Phe Phe Trp Gly Val Gly Leu Leu Ala Ser Leu Leu Thr Leu  
305 310 315 320

60 Ala Ile Leu Tyr Leu Met  
325

WO 98/26072

PCT/US97/22578

-177-

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 369 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 1..366  
 (D) OTHER INFORMATION: Dpj

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

25	ATG AGA ATG ATA GTT GGA CAC GGA ATT GAC ATC GAA GAA TTG GCT TCG	48
	Met Arg Met Ile Val Gly His Gly Ile Asp Ile Glu Glu Leu Ala Ser	
	1 5 10 15	
30	ATA GAA AGC GCA GTT ACA CGA CAT GAA GGA TTT GCT AAG CGT GTA CTG	96
	Ile Glu Ser Ala Val Thr Arg His Glu Gly Phe Ala Lys Arg Val Leu	
	20 25 30	
35	ACC GCT CAG GAA ATG GAG CGC TTC ACC AGT CTC AAA GGA CGC AGG CAA	144
	Thr Ala Gln Glu Met Glu Arg Phe Thr Ser Leu Lys Gly Arg Arg Gln	
	35 40 45	
40	ATA GAA TAT TTA GCT GGT CGC TGG TCG GCT AAG GAG GCC TTT TCC AAG	192
	Ile Glu Tyr Leu Ala Gly Arg Trp Ser Ala Lys Glu Ala Phe Ser Lys	
	50 55 60	
45	GCT ATG GGA ACG GGC ATT AGC AAG CTC GGT TTT CAG GAT TTG GAA GTC	240
	Ala Met Gly Thr Gly Ile Ser Lys Leu Gly Phe Gln Asp Leu Glu Val	
	65 70 75 80	
50	TTG AAC AAT GAA CGT GGG GCG CCT TAT TTT AGT CAG GCA CCA TTT TCA	288
	Leu Asn Asn Glu Arg Gly Ala Pro Tyr Phe Ser Gln Ala Pro Phe Ser	
	85 90 95	
55	GGA AAG ATT TGG CTG TCT ATC AGC CAC ACC GAT CAG TTT GTG ACA GCC	336
	Gly Lys Ile Trp Leu Ser Ile Ser His Thr Asp Gln Phe Val Thr Ala	
	100 105 110	
60	AGT GTC ATT TTG GAG GAA AAT CAT GAA AGC TAG	369
	Ser Val Ile Leu Glu Glu Asn His Glu Ser	
	115 120	

(2) INFORMATION FOR SEQ ID NO:102:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 122 amino acids

PCT/US97/22578

(B) TYPE: amino acid  
(D) TOPOLOGY: linear

67

10

10

10

15

20

20

25

25

30

35

35

40

45

55

60

ACT ATT AGT GGT GCT AAA AAT AGT GTC GTT GCC TTA ATT CCA GCT ATT 96  
Thr Ile Ser Gly Ala Lys Asn Ser Val Val Ala Leu Ile Pro Ala Ile



W09826072

PCT/US97/22578

-179-

	20	25	30	
5	ATC TTG GCT GAT GAT GTG GTG ACT TTG GAT TGC GTT CCA GAT ATT TCG Ile Leu Ala Asp Asp Val Val Thr Leu Asp Cys Val Pro Asp Ile Ser 35 40 45	144		
10	GAT GTA GCC AGT CTT GTC GAA ATC ATG GAA ATG GGA GCT ACT GTT Asp Val Ala Ser Leu Val Glu Ile Met Glu Leu Met 50 55 60	192		
15	AAG CGT TAT GAC GAT GTA TTG GAG ATT GAC CCA AGA GGT GTT CAA AAT Lys Arg Tyr Asp Asp Val Leu Glu Ile Asp Pro Arg Gly Val Gln Asn 65 70 75 80	240		
20	ATT CCA ATG CCT TAT GGT AAA ATT AAC AGT CTT CGT GCA TCT TAC TAT Ile Pro Met Pro Tyr Gly Lys Ile Asn Ser Leu Arg Ala Ser Tyr Tyr 85 90 95	288		
25	TTT TAT GGG AGC CTC TTA GGC CGT TTT GGT GAA GCG ACA GTT GGT CTA Phe Tyr Gly Ser Leu Leu Gly Arg Phe Gly Glu Ala Thr Val Gly Leu 100 105 110	336		
30	CCG GGA GGA TGT GAT CTT GGT CCT CGT CCG ATT GAC TTA CAC CTT AAG Pro Gly Gly Cys Asp Leu Gly Pro Arg Pro Ile Asp Leu His Leu Lys 115 120 125	384		
35	GCG TTT GAA GCT ATG GGT GCC ACT GCT AGC TAC GAG GGA GAT AAC ATG Ala Phe Glu Ala Met Gly Ala Thr Ala Ser Tyr Glu Gly Asp Asn Met 130 135 140	432		
40	AAG TTA TCT GCT AAA GAT ACA GGA CTT CAT GGT GCA ACT ATT TAC ATG Lys Leu Ser Ala Lys Asp Thr Gly Leu His Gly Ala Ser Ile Tyr Met 145 150 155	480		
45	GAT ACG GTT AGT GTG GGA GCA ACG ATT AAT ACG ATG ATT GCT GCG GTT Asp Thr Val Ser Val Gly Ala Thr Ile Asn Thr Met Ile Ala Ala Val 165 170 175	528		
50	AAA GCA AAT GGT CGT ACT ATT ATT GAA AAT GCA GCC CGT GAA CCT GAG Lys Ala Asn Gly Arg Thr Ile Ile Glu Asn Ala Ala Arg Glu Pro Glu 180 185 190	576		
55	ATT ATT CAT GTA GCT ACT CTC TTG AAT AAT ATG GGT GCC CAT ATC CGT Ile Ile Asp Val Ala Thr Leu Leu Asn Asn Met Gly Ala His Ile Arg 195 200 205	624		
60	GGG GCA GGA ACT AAT ATC ATC ATT ATT GAT GGT GTT GAA AGA TTA CAT Gly Ala Gly Thr Asn Ile Ile Ile Ile Asp Gly Val Glu Arg Leu His 210 215 220	672		
65	GGG ACA CGT CAT CAG GTG ATT CCA GAC CGC ATT GAA GCT GGA ACA TAT Gly Thr Arg His Gln Val Ile Pro Asp Arg Ile Glu Ala Gly Thr Tyr 225 230 235 240	720		
70	ATA TCT TTA GCT GCT GCA GTT GGT AAA GGA ATT CCT ATA AAT AAT GTT Ile Ser Leu Ala Ala Val Gly Lys Gly Ile Arg Ile Asn Asn Val 245 250 255	768		
75	CTT TAC GAA CAC CTG GAA GGG TTT GTT GCT AAG TTG GAA GAA ATG GGA Leu Tyr Glu His Leu Glu Gly Phe Val Ala Lys Leu Glu Glu Met Gly 260 265 270	816		

WO 98/26072

PCT/US97/22578

-180-

GTG AGA ATG ACT GTA TCT GAA GAC AGC ATT TTT GTC GAG GAA CAG TCT 864  
Val Arg Met Thr Val Ser Glu Asp Ser Ile Phe Val Glu Glu Gln Ser  
275 280 285

5 AAT TTG AAA GCA ATC AAT ATT AAG ACA GCT CCT TAC CCA GGC TTT GCA 912  
Asn Leu Lys Ala Ile Asn Ile Lys Thr Ala Pro Tyr Pro Gly Phe Ala  
290 295 300

10 ACT GAT TTG CAA CAA CCG CTT ACC CCT CTT TTA CTA AGA GCG AAT GGT 960  
Thr Asp Leu Gln Gln Pro Leu Thr Pro Leu Leu Arg Ala Asn Gly  
305 310 315 320

15 CGT GGT ACA ATT GTC GAT ACG ATT TAC GAA AAA CGT GTA AAT CAT GTT 1008  
Arg Gly Thr Ile Val Asp Thr Ile Tyr Glu Lys Arg Val Asn His Val  
325 330 335

20 TTT GAA CTA GCA AAG ATG GAT GCG GAT ATT TCG ACA ACA AAT GGT CAT 1056  
Phe Glu Leu Ala Lys Met Asp Ala Asp Ile Ser Thr Thr Asn Gly His  
340 345 350

25 ATT TTG TAC ACG GGT GGA CGT GAT TTA CGT GGT GCC AGT GTT AAA GCG 1104  
Ile Leu Tyr Thr Gly Gly Arg Asp Leu Arg Gly Ala Ser Val Lys Ala  
355 360 365

30 ACC GAC TTA AGA GCT GGG GCT GCA CTA GTC ATT GCT GGG CTT ATG GCT 1152  
Thr Asp Leu Arg Ala Gly Ala Ala Leu Val Ile Ala Gly Leu Met Ala  
370 375 380

35 GAA GGC AAA ACT GAA ATT ACC AAT ATC GAG TTT ATC TTA CGT GGT TAT 1200  
Glu Gly Lys Thr Glu Ile Thr Asn Ile Glu Phe Ile Leu Arg Gly Tyr  
385 390 395 400

40 TCT GAT ATT ATC GAA AAA TTA CGT AAT TTA GGA GCG GAT ATT AGA CTT 1248  
Ser Asp Ile Ile Glu Lys Leu Arg Asn Leu Gly Ala Asp Ile Arg Leu  
405 410 415

GTT GAG GAT TAA 1260  
Val Glu Asp  
419

(2) INFORMATION FOR SEQ ID NO:104:

45 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 419 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

55 Met Arg Lys Ile Val Ile Asn Gly Gly Leu Pro Leu Gln Gly Glu Ile  
1 5 10 15

Thr Ile Ser Gly Ala Lys Asn Ser Val Val Ala Leu Ile Pro Ala Ile  
20 25 30

60 Ile Leu Ala Asp Asp Val Val Thr Leu Asp Cys Val Pro Asp Ile Ser  
35 40 45

WO 98/26072

PCT/US97/22578

-181-

Asp Val Ala Ser Leu Val Glu Ile Met Glu Leu Met Gly Ala Thr Val  
50 55 60

5 Lys Arg Tyr Asp Asp Val Leu Glu Ile Asp Pro Arg Gly Val Gln Asn  
65 70 75 80

Ile Pro Met Pro Tyr Gly Lys Ile Asn Ser Leu Arg Ala Ser Tyr Tyr  
85 90 95

10 Phe Tyr Gly Ser Leu Leu Gly Arg Phe Gly Glu Ala Thr Val Gly Leu  
100 105 110

15 Pro Gly Gly Cys Asp Leu Gly Pro Arg Pro Ile Asp Leu His Leu Lys  
115 120 125

Ala Phe Glu Ala Met Gly Ala Thr Ala Ser Tyr Glu Gly Asp Asn Met  
130 135 140

20 Lys Leu Ser Ala Lys Asp Thr Gly Leu His Gly Ala Ser Ile Tyr Met  
145 150 155 160

Asp Thr Val Ser Val Gly Ala Thr Ile Asn Thr Met Ile Ala Ala Val  
165 170 175

25 Lys Ala Asn Gly Arg Thr Ile Ile Glu Asn Ala Ala Arg Glu Pro Glu  
180 185 190

30 Ile Ile Asp Val Ala Thr Leu Leu Asn Asn Met Gly Ala His Ile Arg  
195 200 205

Gly Ala Gly Thr Asn Ile Ile Ile Ile Asp Gly Val Glu Arg Leu His  
210 215 220

35 Gly Thr Arg His Gln Val Ile Pro Asp Arg Ile Glu Ala Gly Thr Tyr  
225 230 235 240

Ile Ser Leu Ala Ala Ala Val Gly Lys Gly Ile Arg Ile Asn Asn Val  
245 250 255

40 Leu Tyr Glu His Leu Glu Gly Phe Val Ala Lys Leu Glu Glu Met Gly  
260 265 270

45 Val Arg Met Thr Val Ser Glu Asp Ser Ile Phe Val Glu Glu Gln Ser  
275 280 285

Asn Leu Lys Ala Ile Asn Ile Lys Thr Ala Pro Tyr Pro Gly Phe Ala  
290 295 300 305

50 Thr Asp Leu Gln Gln Pro Leu Thr Pro Leu Leu Leu Arg Ala Asn Gly  
305 310 315 320

Arg Gly Thr Ile Val Asp Thr Ile Tyr Glu Lys Arg Val Asn His Val  
325 330 335

55 Phe Glu Leu Ala Lys Met Asp Ala Asp Ile Ser Thr Thr Asn Gly His  
340 345 350

60 Ile Leu Tyr Thr Gly Gly Arg Asp Leu Arg Gly Ala Ser Val Lys Ala  
355 360 365



WO 98/26072

PCT/US97/22578

-183-

	Ala	Asp	Met	Val	Phe	Ile	Thr	Ala	Gly	Met	Gly	Gly	Gly	Ser	Gly	Thr	
				100					105					110			
5	GGA	GCT	GCT	CCT	GTT	ATT	GCT	CGT	ATC	GCC	AAA	GAT	TTA	GGT	CGC	CTT	384
	Gly	Ala	Ala	Pro	Val	Ile	Ala	Arg	Ile	Ala	Lys	Asp	Leu	Gly	Ala	Leu	
				115				120					125				
10	ACA	GTT	GGT	GTT	GTA	ACA	CGT	CCC	TTT	GGT	TTT	GAA	GGA	AGT	AAG	CGT	432
	Thr	Val	Gly	Val	Val	Thr	Arg	Pro	Phe	Gly	Phe	Glu	Gly	Ser	Lys	Arg	
				130			135					140					
15	GGA	CAA	TTT	GCT	GTA	GAA	GGA	ATC	AAT	CAA	CTT	CGT	GAG	CAT	GTA	GAC	480
	Gly	Gln	Phe	Ala	Val	Glu	Gly	Ile	Asn	Gln	Leu	Arg	Glu	His	Val	Asp	
						150					155				160		
20	ACT	CTA	TTG	ATT	ATC	TCA	AAC	AAC	AAT	TTG	CTT	GAA	ATT	GTT	GAT	AAG	528
	Thr	Leu	Leu	Ile	Ile	Ser	Asn	Asn	Asn	Leu	Leu	Glu	Ile	Val	Asp	Lys	
						165				170					175		
25	AAA	ACA	CCG	CTT	TTG	GAG	GCT	CTT	AGC	GAA	GCG	GAT	AAC	GTT	CTT	CGT	576
	Lys	Thr	Pro	Leu	Leu	Glu	Ala	Leu	Ser	Glu	Ala	Asp	Asn	Val	Leu	Arg	
						180			185					190			
30	CAA	GGT	GTT	CAA	GGG	ATT	ACC	GAT	TTG	ATT	ACC	AAT	CCA	GGA	TTG	ATT	624
	Gln	Gly	Val	Gln	Gly	Ile	Thr	Asp	Leu	Ile	Thr	Asn	Pro	Gly	Leu	Ile	
				195				200					205				
35	AAC	CTT	GAC	TTT	GCC	GAT	GTG	AAA	ACG	GTA	ATG	GCA	AAC	AAA	GGG	AAT	672
	Asn	Leu	Asp	Phe	Ala	Asp	Val	Lys	Thr	Val	Met	Ala	Asn	Lys	Gly	Asn	
				210			215					220					
40	GCT	CTT	ATG	GGT	ATT	GGT	ATC	GGT	AGT	GGA	GAA	GAA	CGT	GTG	GTA	GAA	720
	Ala	Leu	Met	Gly	Ile	Gly	Ile	Gly	Ser	Gly	Glu	Glu	Arg	Val	Val	Glu	
						225	230			235					240		
45	GCG	GCA	CGT	AAG	GCA	ATC	TAT	TCA	CCA	CTT	CTT	GAA	ACA	ACT	ATT	GAC	768
	Ala	Ala	Arg	Lys	Ala	Ile	Tyr	Ser	Pro	Leu	Leu	Glu	Thr	Thr	Ile	Asp	
						245			250						255		
50	GGT	GCT	GAG	GAT	GTT	ATC	GTC	AAC	GTT	ACT	GGT	GGT	CTT	GAC	TTA	ACC	816
	Gly	Ala	Glu	Asp	Val	Ile	Val	Asn	Val	Thr	Gly	Gly	Leu	Asp	Leu	Thr	
						260			265					270			
55	TTG	ATT	GAG	GCA	GAA	GAG	GCT	TCA	CAA	ATT	GTG	AAC	CAG	GCA	GCA	GGT	864
	Leu	Ile	Glu	Ala	Glu	Glu	Ala	Ser	Gln	Ile	Val	Asn	Gln	Ala	Ala	Gly	
				275			280						285				
60	CAA	GGA	GTG	AAC	ATC	TGG	CTC	GGT	ACT	TCA	ATT	GAT	GAA	AGT	ATG	CGT	912
	Gln	Gly	Val	Asn	Ile	Trp	Leu	Gly	Thr	Ser	Ile	Asp	Glu	Ser	Met	Arg	
				290			295					300					
65	GAT	GAA	ATT	CGT	GTA	ACA	GTT	GTC	GCA	ACG	GGT	GTT	CGT	CAA	GAC	CGC	960
	Asp	Glu	Ile	Arg	Val	Thr	Val	Val	Ala	Thr	Gly	Val	Arg	Gln	Asp	Arg	
				305			310				315				320		
70	GTA	GAA	AAG	GTT	GTG	GCT	CCA	CAA	GCT	AGA	TCA	CCG	CGC	CTA	GGA	TAA	1008
	Val	Glu	Lys	Val	Val	Ala	Pro	Gln	Ala	Arg	Ser	Pro	Arg	Leu	Gly	*	
						325			330						335		

(2) INFORMATION FOR SEQ ID NO:106:

WO 98/26072

PCT/US97/22578

-184-

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 336 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

Met Thr Phe Ser Phe Asp Thr Ala Ala Ala Gln Gly Ala Val Ile Lys  
1 5 10 15  
Val Ile Gly Val Gly Gly Gly Gly Asn Ala Ile Asn Arg Met Val  
15 20 25 30  
Asp Glu Gly Val Thr Gly Val Glu Phe Ile Ala Ala Asn Thr Asp Val  
35 40 45  
Gln Ala Leu Ser Ser Thr Lys Ala Glu Thr Val Ile Gln Leu Gly Pro  
50 55 60  
Lys Leu Thr Arg Gly Leu Gly Ala Gly Gly Gln Pro Glu Val Gly Arg  
65 70 75 80  
Lys Ala Ala Glu Glu Ser Glu Glu Thr Leu Thr Glu Ala Ile Ser Gly  
85 90 95  
Ala Asp Met Val Phe Ile Thr Ala Gly Met Gly Gly Gly Ser Gly Thr  
100 105 110  
Gly Ala Ala Pro Val Ile Ala Arg Ile Ala Lys Asp Leu Gly Ala Leu  
115 120 125  
Thr Val Gly Val Val Thr Arg Pro Phe Gly Phe Glu Gly Ser Lys Arg  
130 135 140  
Gly Gln Phe Ala Val Glu Gly Ile Asn Gln Leu Arg Glu His Val Asp  
145 150 155 160  
Thr Leu Leu Ile Ile Ser Asn Asn Asn Leu Leu Glu Ile Val Asp Lys  
165 170 175  
Lys Thr Pro Leu Leu Glu Ala Leu Ser Glu Ala Asp Asn Val Leu Arg  
180 185 190  
Gln Gly Val Gln Gly Ile Thr Asp Leu Ile Thr Asn Pro Gly Leu Ile  
195 200 205  
Asn Leu Asp Phe Ala Asp Val Lys Thr Val Met Ala Asn Lys Gly Asn  
210 215 220  
Ala Leu Met Gly Ile Gly Ile Gly Ser Gly Glu Glu Arg Val Val Glu  
225 230 235 240  
Ala Ala Arg Lys Ala Ile Tyr Ser Pro Leu Leu Glu Thr Thr Ile Asp  
245 250 255  
Gly Ala Glu Asp Val Ile Val Asn Val Thr Gly Gly Leu Asp Leu Thr  
260 265 270

-165-

Leu Ile Glu Ala Glu Glu Ala Ser Gln Ile Val Asn Gln Ala Ala Gly  
 275 280 285  
 5 Gln Gly Val Asn Ile Trp Leu Gly Thr Ser Ile Asp Glu Ser Met Arg  
 290 295 300  
 Asp Glu Ile Arg Val Thr Val Val Ala Thr Gly Val Arg Gln Asp Arg  
 305 310 315 320  
 10 Val Glu Lys Val Val Ala Pro Gln Ala Arg Ser Pro Arg Leu Gly  
 325 330 335

(2) INFORMATION FOR SEQ ID NO:107:  
 15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 525 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 20 (ii) MOLECULE TYPE: DNA (genomic)  
 (iii) HYPOTHETICAL: NO  
 25 (iv) ANTI-SENSE: NO  
 (ix) FEATURE:  
 30 (A) NAME/KEY: CDS  
 (B) LOCATION: 1..525  
 (D) OTHER INFORMATION: grpE  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:  
 35 ATG GCC CAA GAT ATA AAA AAT GAA GAA GTA GAA GAA GTT CAA GAA GAG 48  
 Met Ala Gln Asp Ile Lys Asn Glu Glu Val Glu Glu Val Gln Glu Glu  
 1 5 10 15  
 40 GAA GTT GTG GAA ACA GCT GAA GAA ACA ACT CCT GAA AAG TCT GAG TTG 96  
 Glu Val Val Glu Thr Ala Glu Glu Thr Thr Pro Glu Lys Ser Glu Leu  
 20 25 30  
 45 GAC TTG GCA AAT GAA CGT GCA GAT GAG TTC GAA AAC AAA TAT CTT CGC 144  
 Asp Leu Ala Asn Glu Arg Ala Asp Glu Phe Glu Asn Lys Tyr Leu Arg  
 35 40 45  
 50 GCT CAT GCA GAA ATG CAA AAT ATC CAA CGC CGT GCC AAT GAA GAA CGT 192  
 Ala His Ala Glu Met Gln Asn Ile Gln Arg Arg Ala Asn Glu Glu Arg  
 50 55 60  
 55 CAA AAC TTG CAA CGT TAT CGT AGC CAG GAC TTG GCA AAA GCA ATC TTA 240  
 Gln Asn Leu Gln Arg Tyr Arg Ser Gln Asp Leu Ala Lys Ala Ile Leu  
 65 70 75 80  
 60 CCA TCT CTT GAC AAC CTT GAG CGT GCA CTT GCA GTT GAA GGT TTG ACA 288  
 Pro Ser Leu Asp Asn Leu Glu Arg Ala Leu Ala Val Glu Gly Leu Thr  
 85 90 95  
 60 GAT GAT GTG AAG AAG GGC TTG GCG ATG GTG CAA GAA AGC TTG ATT CAC 336  
 Asp Asp Val Lys Lys Gly Leu Ala Met Val Gln Glu Ser Leu Ile His

WO 98/26072

PCT/US97/22578

-186-

	100	105	110	
5	GCT TTG AAA GAA GAA GGA ATT GAA GAA ATC GCA GCA GAT GGC GAA TTT Ala Leu Lys Glu Glu Gly Ile Glu Glu Ile Ala Ala Asp Gly Glu Phe 115	120	125	384
10	GAC CAT AAC TAC CAT ATG GCC ATC CAA ACT CTC CCA GGA GAC GAT GAA Asp His Asn Tyr His Met Ala Ile Gln Thr Leu Pro Gly Asp Asp Glu 130	135	140	432
15	CAC CCA GTA GAT ACC ATC GCC CAA GTC TTT CAA AAA GGC TAC AAA CTC His Pro Val Asp Thr Ile Ala Gln Val Phe Gln Lys Gly Tyr Lys Leu 145	150	155	480
20	CAT GAC CGC ATC CTA CGC CCA GCA ATG GTA GTG GTG TAT AAC TAA His Asp Arg Ile Leu Arg Pro Ala Met Val Val Val Tyr Asn * 165	170	174	525
(2) INFORMATION FOR SEQ ID NO:108:				
(i) SEQUENCE CHARACTERISTICS:				
(A) LENGTH: 174 amino acids				
(B) TYPE: amino acid				
(D) TOPOLOGY: linear				
(ii) MOLECULE TYPE: protein				
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:				
30	Met Ala Gln Asp Ile Lys Asn Glu Glu Val Glu Glu Val Gln Glu Glu 1 5 10 15			
35	Glu Val Val Glu Thr Ala Glu Glu Thr Thr Pro Glu Lys Ser Glu Leu 20 25 30			
40	Asp Leu Ala Asn Glu Arg Ala Asp Glu Phe Glu Asn Lys Tyr Leu Arg 35 40 45			
45	Ala His Ala Glu Met Gln Asn Ile Gln Arg Arg Ala Asn Glu Glu Arg 50 55 60			
50	Gln Asn Leu Gln Arg Tyr Arg Ser Gln Asp Leu Ala Lys Ala Ile Leu 65 70 75 80			
55	Pro Ser Leu Asp Asn Leu Glu Arg Ala Leu Ala Val Glu Gly Leu Thr 85 90 95			
60	Asp Asp Val Lys Lys Gly Leu Ala Met Val Gln Glu Ser Leu Ile His 100 105 110			
65	Ala Leu Lys Glu Glu Gly Ile Glu Glu Ile Ala Ala Asp Gly Glu Phe 115 120 125			
70	Asp His Asn Tyr His Met Ala Ile Gln Thr Leu Pro Gly Asp Asp Glu 130 135 140			
75	His Pro Val Asp Thr Ile Ala Gln Val Phe Gln Lys Gly Tyr Lys Leu 145 150 155 160			
80	His Asp Arg Ile Leu Arg Pro Ala Met Val Val Val Tyr Asn			





WO 98/26072

PCT/US97/22578

- 188 -

[illegible]

(2) INFORMATION FOR SEQ ID NO:110:

```

20      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 194 amino acids
          (B) TYPE: amino acid
          (D) TOPOLOGY: linear

25      (ii) MOLECULE TYPE: protein

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

```

[illegible]

WO 98/26072

PCT/US97/22578

-189-

Asp Asp Val Arg Leu His Gln Tyr Phe Ile Asn Met Cys Lys Glu Lys  
180 185 190

Ser

5

(2) INFORMATION FOR SEQ ID NO:111:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 546 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

15

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

20

(ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 1..543  
(D) OTHER INFORMATION: pgsA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

30	ATG AAA AAA GAA CAA ATT CCC AAT CTC TTA ACA ATA GGT CGA ATT CTC	48
	Met Lys Lys Glu Gln Ile Pro Asn Leu Thr Ile Gly Arg Ile Leu	
	1 5 10	
35	TTT ATA CCT ATT TTT ATC TTT ATT TTA ACG ATA GGA AAT TCG ATA GAG	96
	Phe Ile Pro Ile Phe Ile Phe Ile Leu Thr Ile Gly Asn Ser Ile Glu	
	20 25 30	
40	AGT CAT ATA GTT GCA GCT ATT ATC TTT GCT GTT GCC AGT ATT ACC GAC	144
	Ser His Ile Val Ala Ala Ile Ile Phe Ala Val Ala Ser Ile Thr Asp	
	35 40 45	
45	TAT TTA GAT GGA TAT TTA GCT CGT AAA TCG AAT GTG GTC AGT AAT TTT	192
	Tyr Leu Asp Gly Tyr Leu Ala Arg Lys Trp Asn Val Val Ser Asn Phe	
	50 55 60	
50	GGT AAA TTT GCA GAT CCT ATG GCG GAT AAG TTA CTA GTT ATG TCG GCT	240
	Gly Lys Phe Ala Asp Pro Met Ala Asp Lys Leu Leu Val Met Ser Ala	
	65 70 75 80	
55	TTT ATT ATG TTG ATT GAG TTA GGT ATG GCT CCG GCT TGG ATT GTT GCA	288
	Phe Ile Met Leu Ile Glu Leu Gly Met Ala Pro Ala Trp Ile Val Ala	
	85 90 95	
60	GTG ATT ATC TGT CGT GAG TTA GCT GTG ACA GGT TTA AGG CTT TTA TTG	336
	Val Ile Ile Cys Arg Glu Leu Ala Val Thr Gly Leu Arg Leu Leu Leu	
	100 105 110	
65	GTT GAA ACT GGT GGA ACA ATT TTA GCA GCA GCA ATG CCT GGA AAA ATT	384
	Val Glu Thr Thr Gly Gly Thr Ile Leu Ala Ala Ala Met Pro Gly Lys Ile	
	115 120 125	

WO 98/26072

PCT/US97/22578

-190-

```

AAA ACT TTT AGT CAG ATG TTT GCT ATT ATT TTC TTG CTA TTA CAT TGG      432
Lys Thr Phe Ser Gln Met Phe Ala Ile Ile Phe Leu Leu Leu His Trp
130                      135                      140

5  ACT TTG CTT GGT CAA GTT CTA CTT TAT GTA GCC TTA TTT TTC ACT ATC      480
Thr Leu Leu Gly Gln Val Leu Leu Tyr Val Ala Leu Phe Phe Thr Ile
145                      150                      155

10 TAC TCT GGC TAT GAC TAT TTC AAG GGT AGT GCC TAT GTA CTT AAA GGG      528
Tyr Ser Gly Tyr Asp Tyr Phe Lys Gly Ser Ala Tyr Val Phe Lys Gly
165                      170                      175

ACA TTT GGT TCG AAA TGA      546
Thr Phe Gly Ser Lys
180

(2) INFORMATION FOR SEQ ID NO:112:

20      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 181 amino acids
          (B) TYPE: amino acid
          (D) TOPOLOGY: linear

25      (ii) MOLECULE TYPE: protein

          (xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

30  Met Lys Lys Glu Gln Ile Pro Asn Leu Leu Thr Ile Gly Arg Ile Leu
    1      5      10      15

    Phe Ile Pro Ile Phe Ile Phe Ile Leu Thr Ile Gly Asn Ser Ile Glu
    20      25      30

35  Ser His Ile Val Ala Ala Ile Ile Phe Ala Val Ala Ser Ile Thr Asp
    35      40      45

    Tyr Leu Asp Gly Tyr Leu Ala Arg Lys Trp Asn Val Val Ser Asn Phe
    50      55      60

40  Gly Lys Phe Ala Asp Pro Met Ala Asp Lys Leu Leu Val Met Ser Ala
    65      70      75      80

    Phe Ile Met Leu Ile Glu Leu Gly Met Ala Pro Ala Trp Ile Val Ala
    85      90      95

45  Val Ile Ile Cys Arg Glu Leu Ala Val Thr Gly Leu Arg Leu Leu Leu
    100     105     110

50  Val Glu Thr Gly Gly Thr Ile Leu Ala Ala Ala Met Pro Gly Lys Ile
    115     120     125

    Lys Thr Phe Ser Gln Met Phe Ala Ile Ile Phe Leu Leu Leu His Trp
    130     135     140

55  Thr Leu Leu Gly Gln Val Leu Leu Tyr Val Ala Leu Phe Phe Thr Ile
    145     150     155     160

60  Tyr Ser Gly Tyr Asp Tyr Phe Lys Gly Ser Ala Tyr Val Phe Lys Gly
    165     170     175

```

WO 98/26072

PCT/US97/22578

-191-

Thr Phe Gly Ser Lys  
180

(2) INFORMATION FOR SEQ ID NO:113:

5

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1224 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

15

(iv) ANTI-SENSE: NO

(ix) FEATURE:

20

- (A) NAME/KEY: CDS  
(B) LOCATION: 1..1221  
(D) OTHER INFORMATION: RodA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

	ATG AAA CGT TCT CTC GAC TCT AGA GTC GAT TAT AGT TTG CTC TTG CCA	48
	Met Lys Arg Ser Leu Asp Ser Arg Val Asp Tyr Ser Leu Leu Leu Pro	
	1 5 10 15	
30	CTA TTT TTT CTA CTG GTC ATC GGT GTG GCT ATC TAT ATA GCC GTT	96
	Val Phe Phe Leu Leu Val Ile Gly Val Val Ala Ile Tyr Ile Ala Val	
	20 25 30	
35	AGT CAT GAT TAT CCC AAT AAT ATT CTG CCC ATT TTA GGG CAG CAG GTC	144
	Ser His Asp Tyr Pro Asn Asn Ile Leu Pro Ile Leu Gly Gln Gln Val	
	35 40 45	
40	GCC TGG ATT GCC TTG GGG CTT GTG ATT GGT TTT GTG GTC ATG CTC TTT	192
	Ala Trp Ile Ala Leu Gly Leu Val Ile Gly Phe Val Val Met Leu Phe	
	50 55 60	
45	AAT ACA GAA TTT CTT TGG AAG GTG ACC CCC TTT CTA TAT ATT TTA GGC	240
	Asn Thr Glu Phe Leu Trp Lys Val Thr Pro Phe Leu Tyr Ile Leu Gly	
	65 70 75 80	
50	TTG GGA CTT ATG ATC TTG CCG ATT GTA TTT TAT AAT CCA AGC TTA GTT	288
	Leu Gly Leu Met Ile Leu Pro Ile Val Phe Tyr Asn Pro Ser Leu Val	
	85 90 95	
55	GCA TCA ACG GGT GCC AAA AAC TGG GTA TCA ATA AAT GGA ATT ACC CTA	336
	Ala Ser Thr Gly Ala Lys Asn Trp Val Ser Ile Asn Gly Ile Thr Leu	
	100 105 110	
60	TTT CAA CCG TCA GAA TTT ATG AAG ATA TCC TAT ATC CTC ATG TTG GCT	384
	Phe Gln Pro Ser Glu Phe Met Lys Ile Ser Tyr Ile Leu Met Leu Ala	
	115 120 125	
60	CGT GTC ACT GTC CAA TTT ACA AAG AAA CAT AAG GAA TGG AGA CGC ACG	432
	Arg Val Ile Val Gln Phe Thr Lys Lys His Lys Glu Trp Arg Arg Thr	
	130 135 140	

WO 9826072

PCT/US97/22578

-192-

5	GTT	CCG	CTG	GAC	TTT	TTG	TTA	ATT	TTC	TGG	ATG	ATT	CTC	TTT	ACC	ATT	480
	Val	Pro	Leu	Asp	Phe	Leu	Leu	Ile	Phe	Trp	Met	Ile	Leu	Phe	Thr	Ile	
	145					150					155				160		
10	CCA	GTC	CTA	GTT	CTT	TTA	GCA	CTT	CAA	AGT	GAC	TTG	GGG	ACG	GCT	TTG	528
	Pro	Val	Leu	Val	Leu	Leu	Ala	Leu	Gln	Ser	Asp	Leu	Gly	Thr	Ala	Leu	
						165					170				175		
15	GTT	TTT	GTA	GCC	ATT	TTC	TCA	GGA	ATC	GTT	TTA	TTA	TCA	GGG	GTT	TCT	576
	Val	Phe	Val	Ala	Ile	Phe	Ser	Gly	Ile	Val	Leu	Leu	Ser	Gly	Val	Ser	
					180				185					190			
20	TGG	AAA	ATT	ATT	ATC	CCA	GTA	TTT	CTG	ACT	GCT	GTA	ACA	GGA	GTT	GCT	624
	Trp	Lys	Ile	Ile	Ile	Pro	Val	Phe	Val	Thr	Ala	Val	Thr	Gly	Val	Ala	
			195					200					205				
25	GGT	TTC	TTA	GCT	ATC	TTT	ATT	AGC	AAG	GAC	GGA	CGA	GCT	TTT	CTT	CAC	672
	Gly	Phe	Leu	Ala	Ile	Phe	Ile	Ser	Lys	Asp	Gly	Arg	Ala	Phe	Leu	His	
			210				215					220					
30	CAG	ATT	GGA	ATG	CCG	ACC	TAC	CAA	ATC	AAT	CGG	ATT	TTG	GCT	TGG	CTC	720
	Gln	Ile	Gly	Met	Pro	Thr	Tyr	Gln	Ile	Asn	Arg	Ile	Leu	Ala	Trp	Leu	
			225				230				235					240	
35	AAT	CCC	TTT	GAG	TTT	GCC	CAA	ACA	ACG	ACT	TAC	CAG	CAG	GCT	CAA	GGG	768
	Asn	Pro	Phe	Glu	Phe	Ala	Gln	Thr	Thr	Tyr	Thr	Gln	Gln	Ala	Gln	Gly	
					245					250					255		
40	CAG	ATT	GCC	ATT	GGG	AGT	GGT	GGC	TTA	TTT	GGT	CAG	GGA	TTT	AAT	GCT	816
	Gln	Ile	Ala	Ile	Gly	Ser	Gly	Gly	Leu	Phe	Gly	Gln	Gly	Phe	Asn	Ala	
			260					265					270				
45	TCG	AAT	CTG	CTT	ATC	CCA	GTT	CGA	GAG	TCA	GAT	ATG	ATT	TTT	ACG	GTT	864
	Ser	Asn	Leu	Leu	Ile	Pro	Val	Arg	Glu	Ser	Asp	Met	Ile	Phe	Thr	Val	
			275					280					285				
50	ATT	GCA	GAA	GAT	TTT	GGC	TTT	ATT	GGC	TCT	GTC	CTG	GTT	ATT	GCC	CTC	912
	Ile	Ala	Glu	Asp	Phe	Gly	Phe	Ile	Gly	Ser	Val	Leu	Val	Ile	Ala	Leu	
			290				295					300					
55	TAT	CTC	ATG	TTG	ATT	TAC	CGT	ATG	TTG	AAG	ATT	ACT	CTT	AAA	TCA	AAT	960
	Tyr	Leu	Met	Leu	Ile	Tyr	Arg	Met	Leu	Lys	Ile	Thr	Leu	Lys	Ser	Asn	
			305				310				315				320		
60	AAC	CAG	TTC	TAC	ACT	TAT	ATT	TCC	ACA	GGT	TTG	ATT	ATG	ATG	TTG	CTC	1008
	Asn	Gln	Phe	Tyr	Thr	Tyr	Ile	Ser	Thr	Gly	Leu	Ile	Met	Met	Leu	Leu	
					325						330				335		
65	TTC	CAC	ATC	TTT	GAG	AAT	ATC	GGT	GCT	GTG	ACT	GGA	CTA	CTT	CCT	TTG	1056
	Phe	His	Ile	Phe	Glu	Asn	Ile	Gly	Ala	Val	Thr	Gly	Leu	Leu	Pro	Leu	
					340				345						350		
70	ACG	GGG	ATT	CCC	TTG	CCT	TTC	ATT	TCG	CAA	GGG	GGA	TCA	GCG	ATT	ATC	1104
	Thr	Gly	Ile	Pro	Leu	Pro	Phe	Ile	Ser	Gln	Gly	Gly	Ser	Ala	Ile	Ile	
			355					360					365				
75	AGT	AAT	CTG	ATT	GGT	GTT	GGT	TTG	CTT	TTA	TCG	ATG	AGT	TAC	CAG	ACT	1152
	Ser	Asn	Leu	Ile	Gly	Val	Gly	Leu	Leu	Leu	Ser	Met	Ser	Tyr	Gln	Thr	
			370				375						380				

WO 98/26072

PCT/US97/22578

-193-

AAT CTA GCT GAA GAA AAG AGC GGA AAA GTC CCA TTC AAA CGG AAA AAG 1200  
 Asn Leu Ala Glu Glu Lys Ser Gly Lys Val Pro Phe Lys Arg Lys Lys  
 385 390 395 400

5 GTT GTA TTA AAA CAA ATT AAA TAA 1224  
 Val Val Leu Lys Gln Ile Lys  
 405

10 (2) INFORMATION FOR SEQ ID NO:114:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 407 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: protein  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

20 Met Lys Arg Ser Leu Asp Ser Arg Val Asp Tyr Ser Leu Leu Leu Pro  
 1 5 10 15  
 Val Phe Phe Leu Leu Val Ile Gly Val Val Ala Ile Tyr Ile Ala Val  
 25 20 25 30  
 Ser His Asp Tyr Pro Asn Asn Ile Leu Pro Ile Leu Gly Gln Gln Val  
 35 40 45  
 30 Ala Trp Ile Ala Leu Gly Leu Val Ile Gly Phe Val Val Met Leu Phe  
 50 55 60  
 Asn Thr Glu Phe Leu Trp Lys Val Thr Pro Phe Leu Tyr Ile Leu Gly  
 65 70 75 80  
 35 Leu Gly Leu Met Ile Leu Pro Ile Val Phe Tyr Asn Pro Ser Leu Val  
 85 90 95  
 40 Ala Ser Thr Gly Ala Lys Asn Trp Val Ser Ile Asn Gly Ile Thr Leu  
 100 105 110  
 Phe Gln Pro Ser Glu Phe Met Lys Ile Ser Tyr Ile Leu Met Leu Ala  
 115 120 125  
 45 Arg Val Ile Val Gln Phe Thr Lys Lys His Lys Glu Trp Arg Arg Thr  
 130 135 140  
 Val Pro Leu Asp Phe Leu Leu Ile Phe Trp Met Ile Leu Phe Thr Ile  
 145 150 155 160  
 50 Pro Val Leu Val Leu Leu Ala Leu Gln Ser Asp Leu Gly Thr Ala Leu  
 165 170 175  
 55 Val Phe Val Ala Ile Phe Ser Gly Ile Val Leu Leu Ser Gly Val Ser  
 180 185 190  
 Trp Lys Ile Ile Ile Pro Val Phe Val Thr Ala Val Thr Gly Val Ala  
 195 200 205  
 60 Gly Phe Leu Ala Ile Phe Ile Ser Lys Asp Gly Arg Ala Phe Leu His  
 210 215 220

WO 98/26072

PCT/US97/22578

-194-

Gln Ile Gly Met Pro Thr Tyr Gln Ile Asn Arg Ile Leu Ala Trp Leu  
225 230 235 240

5 Asn Pro Phe Glu Phe Ala Gln Thr Thr Tyr Gln Gln Ala Gln Gly  
245 250 255

Gln Ile Ala Ile Gly Ser Gly Gly Leu Phe Gly Gln Gly Phe Asn Ala  
260 265 270

10 Ser Asn Leu Leu Ile Pro Val Arg Glu Ser Asp Met Ile Phe Thr Val  
275 280 285

15 Ile Ala Glu Asp Phe Gly Phe Ile Gly Ser Val Leu Val Ile Ala Leu  
290 295 300

Tyr Leu Met Leu Ile Tyr Arg Met Leu Lys Ile Thr Leu Lys Ser Asn  
305 310 315 320

20 Asn Gln Phe Tyr Thr Tyr Ile Ser Thr Gly Leu Ile Met Met Leu Leu  
325 330 335

Phe His Ile Phe Glu Asn Ile Gly Ala Val Thr Gly Leu Leu Pro Leu  
340 345 350

25 Thr Gly Ile Pro Leu Pro Phe Ile Ser Gln Gly Gly Ser Ala Ile Ile  
355 360 365

30 Ser Asn Leu Ile Gly Val Gly Leu Leu Ser Met Ser Tyr Gln Thr  
370 375 380

Asn Leu Ala Glu Glu Lys Ser Gly Lys Val Pro Phe Lys Arg Lys Lys  
385 390 395 400

35 Val Val Leu Lys Gln Ile Lys  
405

(2) INFORMATION FOR SEQ ID NO:115:

40 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1311 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

50 (iv) ANTI-SENSE: NO

(ix) FEATURE:

55 (A) NAME/KEY: CDS  
(B) LOCATION: 1..1311  
(D) OTHER INFORMATION: SecY

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

60 ATG TTT TTT AAA TTA TTA AGA GAA GCT CTT AAA GTC AAG CAG GTT CGA

48



WO 98/26072

PCT/US97/22578

-195-

	Met	Phe	Phe	Lys	Leu	Leu	Arg	Glu	Ala	Leu	Lys	Val	Lys	Gln	Val	Arg	
	1				5					10					15		
5	TCA	AAA	ATT	TTA	TTT	ACA	ATT	TTT	ATC	GTT	TTG	GTC	TTT	CGT	ATC	GGA	96
	Ser	Lys	Ile	Leu	Phe	Thr	Ile	Phe	Ile	Val	Leu	Val	Phe	Arg	Ile	Gly	
				20					25					30			
10	ACT	AGC	ATT	ACA	GTT	CCT	GGT	GTG	AAT	GCC	AAT	AGC	TTG	AAT	GCT	TTA	144
	Thr	Ser	Ile	Thr	Val	Pro	Gly	Val	Asn	Ala	Asn	Ser	Leu	Asn	Ala	Leu	
				35				40					45				
15	AGT	GGA	TTA	TCC	TTC	TTA	AAC	ATG	TTG	AGC	TTG	GTG	TCG	GGG	AAT	GCC	192
	Ser	Gly	Leu	Ser	Phe	Leu	Asn	Met	Leu	Ser	Leu	Val	Ser	Gly	Asn	Ala	
		50					55					60					
20	CTA	AAA	AAC	TTT	TCG	ATT	TTT	GCC	CTA	GGA	GTT	AGT	CCC	TAT	ATC	ACC	240
	Leu	Lys	Asn	Phe	Ser	Ile	Phe	Ala	Leu	Gly	Val	Ser	Pro	Tyr	Ile	Thr	
		65			70						75			80			
25	GCT	TCT	ATT	GTT	GTG	CAA	CTC	TTG	CAA	ATG	GAT	ATT	TTA	CCC	AAG	TTT	288
	Ala	Ser	Ile	Val	Val	Gln	Leu	Leu	Gln	Met	Asp	Ile	Leu	Pro	Lys	Phe	
					85					90				95			
30	GTA	GAG	TGG	GGT	AAA	CAA	GGG	GAA	GTA	GGT	CGA	AGA	AAA	TTG	AAT	CAA	336
	Val	Glu	Trp	Gly	Lys	Gln	Gly	Glu	Val	Gly	Arg	Arg	Lys	Leu	Asn	Gln	
				100				105						110			
35	GCT	ACT	CGT	TAT	ATT	GCT	CTA	GTT	CTC	GCT	TTT	GTG	CAA	TCT	ATC	GGG	384
	Ala	Thr	Arg	Tyr	Ile	Ala	Leu	Val	Leu	Ala	Phe	Val	Gln	Ser	Ile	Gly	
			115				120						125				
40	ATT	ACA	GCT	GGT	TTT	AAT	ACC	TTG	GCT	GGA	GCT	CAA	TTG	ATT	AAA	ACT	432
	Ile	Thr	Ala	Gly	Phe	Asn	Thr	Leu	Ala	Gly	Ala	Gln	Leu	Ile	Lys	Thr	
		130				135						140					
45	GCT	TTA	ACT	CCA	CAA	GTT	TTT	CTG	ACG	ATT	GGT	ATC	ATC	TTA	ACA	GCT	480
	Ala	Leu	Thr	Pro	Gln	Val	Phe	Leu	Thr	Ile	Gly	Ile	Ile	Leu	Thr	Ala	
		145				150					155			160			
50	GGT	AGT	ATG	ATT	GTG	ACT	TGG	TTG	GGT	GAG	CAA	ATT	ACA	GAT	AAG	GGA	528
	Gly	Ser	Met	Ile	Val	Thr	Trp	Leu	Gly	Glu	Gln	Ile	Thr	Asp	Lys	Gly	
				165					170					175			
55	TAC	GGA	AAC	GGT	GTT	TCC	ATG	ATT	ATC	TTT	GCC	GGG	ATT	GTT	TCC	TCA	576
	Tyr	Gly	Asn	Gly	Val	Ser	Met	Ile	Ile	Phe	Ala	Gly	Ile	Val	Ser	Ser	
				180				185						190			
60	ATT	CCA	GAG	ATG	ATT	CAG	GGC	ATC	TAT	GTG	GAC	TAC	TTT	GTG	AAC	GTC	624
	Ile	Pro	Glu	Met	Ile	Gln	Gly	Ile	Tyr	Val	Asp	Tyr	Phe	Val	Asn	Val	
			195				200						205				
65	CCA	AGT	AGC	CGT	ATC	ACT	TCA	TCT	ATC	ATT	TTC	GTA	ATC	ATT	TTG	ATT	672
	Pro	Ser	Ser	Arg	Ile	Thr	Ser	Ser	Ile	Ile	Phe	Val	Ile	Ile	Leu	Ile	
			210				215					220					
70	ATT	ACT	GTA	TTG	TTG	ATT	TAC	TTT	ACA	ACT	TAT	GTT	CAA	CAA	GCA		720
	Ile	Thr	Val	Leu	Leu	Ile	Ile	Tyr	Phe	Thr	Thr	Tyr	Val	Gln	Gln	Ala	
		225				230				235				240			
75	GAA	TAC	AAA	ATT	CCA	ATC	CAA	TAT	ACT	AAG	GTT	GCA	CAA	GGT	GCT	CCA	768
	Glu	Tyr	Lys	Ile	Pro	Ile	Gln	Tyr	Thr	Lys	Val	Ala	Gln	Gly	Ala	Pro	

WO 98/26072

PCT/US97/22578

-196-

	245	250	255	
5	TCT AGC TCT TAC CTT CCG TTA AAG GTA AAT CCT GCT GGA GTT ATC CCT Ser Ser Ser Tyr Leu Pro Leu Lys Val Asn Pro Ala Gly Val Ile Pro 260 265 270			816
10	GTT ATC TTT GCC AGT TCG ATT ACT GCA GCG CCT GCG GCT ATT CTT CAG Val Ile Phe Ala Ser Ser Ile Thr Thr Ala Ala Pro Ala Ala Ile Leu Gln 275 280			864
15	TTT TTG AGT GCC ACA GGT CAT GAT TGG GCT TGG GTA AGG GTA GCA CAA Phe Leu Ser Ala Thr Gly His Asp Trp Ala Trp Val Arg Val Ala Gln 290 295 300			912
20	GAG ATG TTG GCA ACT ACT TCT CCA ACT GGT ATT GCC ATG TAT GCT TTG Glu Met Leu Ala Thr Thr Ser Pro Thr Gly Ile Ala Met Tyr Ala Leu 305 310 315 320			960
25	TTG ATT ATT CTC TTT ACA TTC TTC TAT ACG TTT GTA CAG ATT AAT CCT Leu Ile Ile Leu Phe Thr Phe Phe Tyr Thr Phe Val Gln Ile Asn Pro 325 330 335			1008
30	GAA AAA GCA GCA GAG AGC CTA CAA AAG AGT GGT GCC TAT ATC CAT GGA Glu Lys Ala Ala Glu Ser Leu Gln Lys Ser Gly Ala Tyr Ile His Gly 340 345 350			1056
35	GTT CGT CCT GGT AAA GGT ACA GAA GAA TAT ATG TCT AAA CTT CTT GGT Val Arg Pro Gly Lys Gly Thr Glu Glu Tyr Met Ser Lys Leu Leu Arg 355 360 365			1104
40	CGT CTT GCA ACT GTT GGT TCC CTC TTC CTT GGT GTG ATT TCC ATT TTA Arg Leu Ala Thr Val Gly Ser Leu Phe Leu Gly Val Ile Ser Ile Leu 370 375 380			1152
45	CCG ATT GCA GCT AAA GAT GTA TTT GGT CTT TCT GAT GTT GTT GCC TTT Pro Ile Ala Ala Lys Asp Val Phe Gly Leu Ser Asp Val Val Ala Phe 385 390 395 400			1200
50	GGT GGA ACA AGT CTC TTG ATC ATT ATC TCT ACA GGT ATC GAA GGA ATC Gly Gly Thr Ser Leu Leu Ile Ile Ile Ser Thr Gly Ile Glu Gly Ile 405 410 415			1248
55	AAG CAA TTG GAA GGT TAC CTA TTG AAA CGT AAG TAT GTT GGT TTC ATG Lys Gln Leu Glu Gly Tyr Leu Leu Lys Arg Lys Tyr Val Gly Phe Met 420 425 430			1296
60	GAC AGA ACA GAA TAA Asp Arg Thr Glu * 435			1311

(2) INFORMATION FOR SEQ ID NO:116:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 437 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

WO 98/26072

PCT/US97/22578

-197-

	Met	Phe	Phe	Lys	Leu	Leu	Arg	Glu	Ala	Leu	Lys	Val	Lys	Gln	Val	Arg	15
5	Ser	Lys	Ile	Leu	Phe	Thr	Ile	Phe	Ile	Val	Leu	Val	Phe	Arg	Ile	Gly	20
	Thr	Ser	Ile	Thr	Val	Pro	Gly	Val	Asn	Ala	Asn	Ser	Leu	Asn	Ala	Leu	35
10	Ser	Gly	Leu	Ser	Phe	Leu	Asn	Met	Leu	Ser	Leu	Val	Ser	Gly	Asn	Ala	50
	Leu	Lys	Asn	Phe	Ser	Ile	Phe	Ala	Leu	Gly	Val	Ser	Pro	Tyr	Ile	Thr	65
15	Ala	Ser	Ile	Val	Val	Gln	Leu	Leu	Gln	Met	Asp	Ile	Leu	Pro	Lys	Phe	85
20	Val	Glu	Trp	Gly	Lys	Gln	Gly	Glu	Val	Gly	Arg	Arg	Lys	Leu	Asn	Gln	100
	Ala	Thr	Arg	Tyr	Ile	Ala	Leu	Val	Leu	Ala	Phe	Val	Gln	Ser	Ile	Gly	115
25	Ile	Thr	Ala	Gly	Phe	Asn	Thr	Leu	Ala	Gly	Ala	Gln	Leu	Ile	Lys	Thr	130
30	Ala	Leu	Thr	Pro	Gln	Val	Phe	Leu	Thr	Ile	Gly	Ile	Ile	Leu	Thr	Ala	145
	Gly	Ser	Met	Ile	Val	Thr	Trp	Leu	Gly	Glu	Gln	Ile	Thr	Asp	Lys	Gly	165
35	Tyr	Gly	Asn	Gly	Val	Ser	Met	Ile	Ile	Phe	Ala	Gly	Ile	Val	Ser	Ser	180
	Ile	Pro	Glu	Met	Ile	Gln	Gly	Ile	Tyr	Val	Asp	Tyr	Phe	Val	Asn	Val	195
40	Pro	Ser	Ser	Arg	Ile	Thr	Ser	Ser	Ile	Ile	Phe	Val	Ile	Ile	Leu	Ile	210
45	Ile	Thr	Val	Leu	Leu	Ile	Ile	Tyr	Phe	Thr	Thr	Tyr	Val	Gln	Gln	Ala	225
	Glu	Tyr	Lys	Ile	Pro	Ile	Gln	Tyr	Thr	Lys	Val	Ala	Gln	Gly	Ala	Pro	245
50	Ser	Ser	Ser	Tyr	Leu	Pro	Leu	Lys	Val	Asn	Pro	Ala	Gly	Val	Ile	Pro	260
	Val	Ile	Phe	Ala	Ser	Ser	Ile	Thr	Ala	Ala	Pro	Ala	Ala	Ile	Leu	Gln	275
55	Phe	Leu	Ser	Ala	Thr	Gly	His	Asp	Trp	Ala	Trp	Val	Arg	Val	Ala	Gln	290
60	Glu	Met	Leu	Ala	Thr	Ser	Ser	Pro	Thr	Gly	Ile	Ala	Met	Tyr	Ala	Leu	305

WO 98/26072

PCT/US97/22578

-198-

```

Leu Ile Ile Leu Phe Thr Phe Phe Tyr Thr Phe Val Gln Ile Asn Pro
      325                               330
5  Glu Lys Ala Ala Glu Ser Leu Gln Lys Ser Gly Ala Tyr Ile His Gly
      340                               345                               350
Val Arg Pro Gly Lys Gly Thr Glu Glu Tyr Met Ser Lys Leu Leu Arg
      355                               360                               365
10 Arg Leu Ala Thr Val Gly Ser Leu Phe Leu Gly Val Ile Ser Ile Leu
      370                               375                               380
Pro Ile Ala Ala Lys Asp Val Phe Gly Leu Ser Asp Val Val Ala Phe
      385                               390                               395                               400
15 Gly Gly Thr Ser Leu Leu Ile Ile Ile Ser Thr Gly Ile Glu Gly Ile
      405                               410                               415
Lys Gln Leu Glu Gly Tyr Leu Leu Lys Arg Lys Tyr Val Gly Phe Met
      420                               425                               430
20 Asp Arg Thr Glu
      435

25 (2) INFORMATION FOR SEQ ID NO:117:
      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 1959 base pairs
          (B) TYPE: nucleic acid
30      (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear
      (ii) MOLECULE TYPE: DNA (genomic)
35      (iii) HYPOTHETICAL: NO
      (iv) ANTI-SENSE: NO

40      (ix) FEATURE:
          (A) NAME/KEY: CDS
          (B) LOCATION: 1..1956
          (D) OTHER INFORMATION: FtsH

45      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:
ATG AAA AAA CAA AAT AAT GGT TTA ATT AAA AAT CCT TTT CTA TGG TTA      48
Met Lys Lys Gln Asn Asn Gly Leu Ile Lys Asn Pro Phe Leu Trp Leu
50      1                               5                               10                               15
TTA TTT ATC TTT TTC CTI GTG ACA GGA TTC CAG TAT TTC TAT TCT GGG      96
Leu Phe Ile Phe Phe Leu Val Thr Gly Phe Gln Tyr Phe Tyr Ser Gly
      20                               25                               30
55 AAT AAC TCA GGA GGA ACT CAG CAA ATC AAC TAT ACT GAG TTG GTA CAA      144
Asn Asn Ser Gly Gly Ser Gln Gln Ile Asn Tyr Thr Glu Leu Val Gln
      35                               40                               45
60 GAA ATT ACC GAT GGT AAT GAA AAA GAA TTA ACT TAC CAA CCA AAT GTT      192
Glu Ile Thr Asp Gly Asn Glu Lys Glu Leu Thr Tyr Gln Pro Asn Val

```

WO 98/26072

PCT/US97/22578

-199-

	50	55	60	
	AGT GTT ATC GAA GTT TCT GGT GTC TAT AAA AAT CCT AAA ACA AGT AAA			240
	Ser Val Ile Glu Val Ser Gly Val Tyr Lys Asn Pro Lys Thr Ser Lys			
5	65	70	75	80
	GAA GGA ACA GGT ATT CAG TTT TTC ACG CCA TCT GTT ACT AAG GTA GAG			288
	Glu Gly Thr Gly Ile Gln Phe Phe Thr Pro Ser Val Thr Lys Val Glu			
	85	90	95	
10	AAA TTT ACC AGC ACT ATT CTT CCT GCA GAT ACT ACC GTA TCA GAA TTG			336
	Lys Phe Thr Ser Thr Ile Leu Pro Ala Asp Thr Thr Val Ser Glu Leu			
	100	105	110	
15	CAA AAA CTT GCT ACT GAC CAT AAA GCA GAA GTA ACT GTT AAG CAT GAA			384
	Gln Lys Leu Ala Thr Asp His Lys Ala Glu Val Thr Val Lys His Glu			
	115	120	125	
20	AGT TCA AGT GGT ATA TGG ATT AAT CTA CTC GTA TCC ATT GTG CCA TTT			432
	Ser Ser Ser Gly Ile Trp Ile Asn Leu Leu Val Ser Ile Val Pro Phe			
	130	135	140	
25	GGA ATT CTA TTC TTC TTC CTA TTC TCT ATG ATG GGA AAT ATG GGA GGA			480
	Gly Ile Leu Phe Phe Phe Leu Phe Ser Met Met Gly Asn Met Gly Gly			
	145	150	155	160
	GGC AAT GGC CGT AAT CCA ATG AGT TTT GGA CGT AGT AAG GCT AAA GCA			528
	Gly Asn Gly Arg Asn Pro Met Ser Phe Gly Arg Ser Lys Ala Lys Ala			
	165	170	175	
30	GCA AAT AAA GAA GAT ATT AAA GTA ACA TTT TCA GAT GTT GCT GGA GCT			576
	Ala Asn Lys Glu Asp Ile Lys Val Arg Phe Ser Asp Val Ala Gly Ala			
	180	185	190	
35	GAG GAA GAA AAA CAA GAA CTA GTT GAA GTT GAT GTC TTA AAA GAT			624
	Glu Glu Glu Lys Gln Glu Leu Val Glu Val Val Glu Phe Leu Lys Asp			
	195	200	205	
40	CCA AAA CGA TTC ACA AAA CTT GGA GCC CGT ATT CCA GCA GGT GTT CTT			672
	Pro Lys Arg Phe Thr Lys Leu Gly Ala Arg Ile Pro Ala Gly Val Leu			
	210	215	220	
45	TTG GAG GGA CCT CCG GCG ACA GGT AAG ACT TTG CTT GCT AAG GCA GTC			720
	Leu Glu Gly Pro Pro Gly Thr Gly Lys Thr Leu Ala Lys Ala Val			
	225	230	235	240
	GCT GGA GAA GCA GGT GTT CCA TTC TTT AGT ATC TCA GGT TCT GAC TTT			768
	Ala Gly Glu Ala Gly Val Pro Phe Phe Ser Ile Ser Gly Ser Asp Phe			
	245	250	255	
50	GTA GAA ATG TTT GTC GGA GTT GGA GCT AGT CGT GTT CGC TCT CTT TTT			816
	Val Glu Met Phe Val Gly Val Gly Ala Ser Arg Val Arg Ser Leu Phe			
	260	265	270	
55	GAG GAT GCC AAA AAA GCA GCA CCA GCT ATC ATC TTT ATC GAT CTA AAT			864
	Glu Asp Ala Lys Lys Ala Ala Pro Ala Ile Ile Phe Ile Asp Leu Asn			
	275	280	285	
60	GAT GCT GTT GGA CGT CAA CGT GGA GTC GGT CTC GGC GGA GGT AAT GAC			912
	Asp Ala Val Gly Arg Gln Arg Gly Val Gly Leu Gly Gly Gly Asn Asp			
	290	295	300	

WO 98/26072

PCT/US97/22578

-200-

5	GAA CGT GAA CAA ACC TTG AAC CAA CTT TTG ATT GAG ATG GAT GGT TTT Glu Arg Glu Gln Thr Leu Asn Gln Leu Leu Ile Glu Met Asp Gly Phe 305 310 315 320	960
	GAG GGA AAT GAA GGG ATT ATC GTC ATC GCT CGG ACA AAC CGT TCA GAT Glu Gly Asn Glu Gly Ile Ile Val Ile Ala Ala Thr Asn Arg Ser Asp 325 330 335	1008
10	GTA CTT GAT CCT GCC CTT TTG CGT CCA GGA CGT TTT GAT AGA AAA GTA Val Leu Asp Pro Ala Leu Leu Arg Pro Gly Arg Phe Asp Arg Lys Val 340 345 350	1056
15	TTG GTT GGC CGT CCT GAT GTT AAA GGT CGT GAA GCA ATC TTG AAA GTT Leu Val Gly Arg Pro Asp Val Lys Gly Arg Glu Ala Ile Leu Lys Val 355 360 365	1104
20	CAC GCT AAG AAC AAG CCT TTA GCA GAA GAT CTT GAT TTG AAA TTA GTG His Ala Lys Asn Lys Pro Leu Ala Glu Asp Val Asp Leu Lys Leu Val 370 375 380	1152
25	GCT CAA CAA ACT CCA GGC TTT GTT GGT GCT GAT TTA GAG AAT GTC TTG Ala Gln Gln Thr Pro Gly Phe Val Gly Ala Asp Leu Glu Asn Val Leu 385 390 395 400	1200
	AAT GAA GCA GCT TTA GTT GCT GCT CGT CGC AAT AAA TCG ATA ATT GAT Asn Glu Ala Ala Leu Val Ala Ala Arg Arg Asn Lys Ser Ile Ile Asp 405 410 415	1248
30	GCT TCA GAT ATT GAT GAA GCA GAA GAT AGA GTT ATT GCT GGA CCT TCT Ala Ser Asp Ile Asp Glu Ala Glu Asp Arg Val Ile Ala Gly Pro Ser 420 425 430	1296
35	AAG AAA GAT AAG ACA GTT TCA CAA AAA GAA CGA GAA TTG GTT GCT TAC Lys Lys Asp Lys Thr Val Ser Gln Lys Glu Arg Glu Leu Val Ala Tyr 435 440 445	1344
40	CAT GAG GCA GGA CAT ACC ATT GTT GGT CTA GTC TTG TCG ACT GCT CGC His Glu Ala Gly His Thr Ile Val Gly Leu Val Leu Ser Thr Ala Arg 450 455 460	1392
45	GTG GTC CAT AAG GTT ACA ATT GTA CCA CGC GGC CGT GCA GGC GGA TAC Val Val His Lys Val Thr Ile Val Pro Arg Gly Arg Ala Gly Gly Tyr 465 470 475 480	1440
	ATG ATT GCA CTT CCT AAA GAG GAT CAA ATG CTT CTA TCT AAA GAA GAT Met Ile Ala Leu Pro Lys Glu Asp Gln Met Leu Leu Ser Lys Glu Asp 485 490 495	1488
50	ATG AAA GAG CAA TTG GCT GGC TTA ATG GGT GGA CGT GTA GCT GAA GAA Met Lys Glu Gln Leu Ala Gly Leu Met Gly Gly Arg Val Ala Glu Glu 500 505 510	1536
55	ATT ATC TTT AAT GTC CAA ACT ACA GGA GCT TCA AAC GAC TTT GAA CAA Ile Ile Phe Asn Val Gln Thr Thr Gly Ala Ser Asn Asp Phe Glu Gln 515 520 525	1584
60	GGC ACA CAA ATG GCA CGT GCA ATG GTT ACA GAG TAC GGT ATG AGT GAA Ala Thr Gln Met Ala Arg Ala Met Val Thr Glu Tyr Gly Met Ser Glu 530 535 540	1632

WO 98/26072

PCT/US97/22578

-201-

AAA CTT GGC CCA GTA CAA TAT GAA GGA AAC CAT GCT ATG CTT GGT GCA 1680  
Lys Leu Gly Pro Val Gln Tyr Glu Gly Asn His Ala Met Leu Gly Ala  
545 550 555 560

5 CAG AGT CCT CAA AAA TCA ATT TCA GAA CAA ACA GCT TAT GAA ATT GAT 1728  
Gln Ser Pro Gln Lys Ser Ile Ser Glu Gln Thr Ala Tyr Glu Ile Asp  
565 570 575

10 GAA GAG GTT CGT TCA TTA TTA AAT GAG GCA CGA AAT AAA GCT GCT GAA 1776  
Glu Glu Val Arg Ser Leu Leu Asn Glu Ala Arg Asn Lys Ala Ala Glu  
580 585 590

15 ATT ATT CAG TCA AAT CGT GAA ACT CAC AAG TTA ATT GCA GAA GCA TTA 1824  
Ile Ile Gln Ser Asn Arg Glu Thr His Lys Leu Ile Ala Glu Ala Leu  
595 600 605

20 TTG AAA TAC GAA ACA TTG GAT AGT ACA CAA ATT AAA GCT CTT TAC GAA 1872  
Leu Lys Tyr Glu Thr Leu Asp Ser Thr Gln Ile Lys Ala Leu Tyr Glu  
610 615 620

ACA GGA AAG ATG CCT GAA GCA GTA GAA GAG GAA TCT CAT GCA CTA TCC 1920  
Thr Gly Lys Met Pro Glu Ala Val Glu Glu Glu Ser His Ala Leu Ser  
625 630 635 640

25 TAT GAT GAA GTA AAG TCA AAA ATG AAT GAC GAA AAA TAA 1959  
Tyr Asp Glu Val Lys Ser Lys Met Asn Asp Glu Lys  
645 650

30 (2) INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 652 amino acids

(B) TYPE: amino acid

35 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

40 Met Lys Lys Gln Asn Asn Gly Leu Ile Lys Asn Pro Phe Leu Trp Leu  
1 5 10 15

45 Leu Phe Ile Phe Phe Leu Val Thr Gly Phe Gln Tyr Phe Tyr Ser Gly  
20 25 30

Asn Asn Ser Gly Gly Ser Gln Gln Ile Asn Tyr Thr Glu Leu Val Gln  
35 40 45

50 Glu Ile Thr Asp Gly Asn Glu Lys Glu Leu Thr Tyr Gln Pro Asn Val  
50 55 60

55 Ser Val Ile Glu Val Ser Gly Val Tyr Lys Asn Pro Lys Thr Ser Lys  
65 70 75 80

Glu Gly Thr Gly Ile Gln Phe Phe Thr Pro Ser Val Thr Lys Val Glu  
85 90 95

60 Lys Phe Thr Ser Thr Ile Leu Pro Ala Asp Thr Thr Val Ser Glu Leu  
100 105 110

PCT/US97/22578

	Gln	Lys	Leu	Ala	Thr	Asp	His	Lys	Ala	Glu	Val	Thr	Val	Lys	His	Glu
	115							120					125			
5	Ser	Ser	Ser	Gly	Ile	Trp	Ile	Asn	Leu	Leu	Val	Ser	Ile	Val	Pro	Phe
	130						135					140				
	Gly	Ile	Leu	Phe	Phe	Phe	Leu	Phe	Ser	Met	Met	Gly	Asn	Met	Gly	Gly
	145					150					155				160	
10	Gly	Asn	Gly	Arg	Asn	Pro	Met	Ser	Phe	Gly	Arg	Ser	Lys	Ala	Lys	Ala
					165					170					175	
	Ala	Asn	Lys	Glu	Asp	Ile	Lys	Val	Arg	Phe	Ser	Asp	Val	Ala	Gly	Ala
				180					185					190		
15	Glu	Glu	Glu	Gly	Lys	Gln	Glu	Leu	Val	Glu	Val	Val	Glu	Phe	Leu	Lys
			195					200					205			Asp
	Pro	Lys	Arg	Phe	Thr	Lys	Leu	Gly	Ala	Arg	Ile	Pro	Ala	Gly	Val	Leu
	210						215					220				
	Leu	Glu	Gly	Pro	Pro	Gly	Thr	Gly	Lys	Thr	Leu	Leu	Ala	Lys	Ala	Val
	225					230					235				240	
25	Ala	Gly	Glu	Ala	Gly	Val	Pro	Phe	Phe	Ser	Ile	Ser	Gly	Ser	Asp	Phe
					245					250					255	
	Val	Glu	Met	Phe	Val	Gly	Val	Gly	Ala	Ser	Arg	Val	Arg	Ser	Leu	Phe
				260					265						270	
30	Glu	Asp	Ala	Lys	Lys	Ala	Ala	Pro	Ala	Ile	Ile	Phe	Ile	Asp	Leu	Asn
			275					280					285			
	Asp	Ala	Val	Gly	Arg	Gln	Arg	Gly	Val	Gly	Leu	Gly	Gly	Gly	Asn	Asp
			290				295					300				
	Glu	Arg	Glu	Gln	Thr	Leu	Asn	Gln	Leu	Leu	Ile	Glu	Met	Asp	Gly	Phe
						310					315				320	
40	Glu	Gly	Asn	Glu	Gly	Ile	Ile	Val	Ile	Ala	Ala	Thr	Asn	Arg	Ser	Asp
					325					330					335	
	Val	Leu	Asp	Pro	Ala	Leu	Leu	Arg	Pro	Gly	Arg	Phe	Asp	Arg	Lys	Val
				340					345					350		
45	Leu	Val	Gly	Arg	Pro	Asp	Val	Lys	Gly	Arg	Glu	Ala	Ile	Leu	Lys	Val
			355					360					365			
	His	Ala	Lys	Asn	Lys	Pro	Leu	Ala	Glu	Asp	Val	Asp	Leu	Lys	Leu	Val
			370				375					380				
	Ala	Gln	Gln	Thr	Pro	Gly	Phe	Val	Gly	Ala	Asp	Leu	Glu	Asn	Val	Leu
						390					395				400	
55	Asn	Glu	Ala	Ala	Leu	Val	Ala	Ala	Arg	Arg	Asn	Lys	Ser	Ile	Ile	Asp
					405					410					415	
	Ala	Ser	Asp	Ile	Asp	Glu	Ala	Glu	Asp	Arg	Val	Ile	Ala	Gly	Pro	Ser
					420				425					430		
60	Lys	Lys	Asp	Lys	Thr	Val	Ser	Gln	Lys	Glu	Arg	Glu	Leu	Val	Ala	Tyr



WO 98/26072

PCT/US97/22578

-203-

	435		440		445
	His Glu Ala Gly His Thr Ile Val Gly Leu Val Leu Ser Thr Ala Arg				
	450		455		460
5	Val Val His Lys Val Thr Ile Val Pro Arg Gly Arg Ala Gly Gly Tyr				
	465		470		475
	Met Ile Ala Leu Pro Lys Glu Asp Gln Met Leu Leu Ser Lys Glu Asp				
10			485		490
	Met Lys Glu Gln Leu Ala Gly Leu Met Gly Gly Arg Val Ala Glu Glu				
			500		505
	Ile Ile Phe Asn Val Gln Thr Thr Gly Ala Ser Asn Asp Phe Glu Gln				
			515		520
	Ala Thr Gln Met Ala Arg Ala Met Val Thr Glu Tyr Gly Met Ser Glu				
			535		540
20	Lys Leu Gly Pro Val Gln Tyr Glu Gly Asn His Ala Met Leu Gly Ala				
			550		555
	Gln Ser Pro Gln Lys Ser Ile Ser Glu Gln Thr Ala Tyr Glu Ile Asp				
25			565		570
	Glu Glu Val Arg Ser Leu Leu Asn Glu Ala Arg Asn Lys Ala Ala Glu				
			580		585
	Ile Ile Gln Ser Asn Arg Glu Thr His Lys Leu Ile Ala Glu Ala Leu				
			595		600
	Leu Lys Tyr Glu Thr Leu Asp Ser Thr Gln Ile Lys Ala Leu Tyr Glu				
			610		615
35	Thr Gly Lys Met Pro Glu Ala Val Glu Glu Ser His Ala Leu Ser				
			625		630
	Tyr Asp Glu Val Lys Ser Lys Met Asn Asp Glu Lys				
			645		650
40					

(2) INFORMATION FOR SEQ ID NO:119:

45	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 1278 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: single
	(D) TOPOLOGY: linear
50	(ii) MOLECULE TYPE: DNA (genomic)
	(iii) HYPOTHETICAL: NO
55	(iv) ANTI-SENSE: NO
	(ix) FEATURE:
	(A) NAME/KEY: CDS
	(B) LOCATION: 1..1278
60	(D) OTHER INFORMATION: FtsY

PCT/US97/22578

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

5	ATG GGA TTG TTT GAC CGT CTA TTC GGA AAA AAA GAA GAA CCT AAA ATC Met Gly Leu Phe Asp Arg Leu Phe Gly Lys Lys Glu Glu Pro Lys Ile	48
10	GAA GAA GTT GTA AAA GAA GCT CTG GAA AAT CTT GAT TTG TCT GAA GAT Glu Glu Val Val Lys Glu Ala Leu Glu Asn Leu Asp Leu Ser Glu Asp	96
15	GTT GAT CCT ACC TTC ACA GAA GTT GAG GAA GTT TCT CAG GAA GAA GCA Val Asp Pro Thr Phe Thr Glu Val GAG Glu Val Phe Glu Glu Glu Ala	144
20	GAG GTT GAA ATT GTT GAA CAA GCT GTG TTC CAA GAA GAG GAA ATC CAA Glu Val Glu Ile Val Glu Glu Ala Val Phe Glu Glu Glu Glu Ile Glu	192
25	GAC ACA GTT GAA GAA AGT CTG GAT TTA GAG CCA GTT GTA GAA GTT TCT Asp Thr Val Glu Glu Ser Leu Asp Leu Glu Pro Val Val Glu Val Ser	240
30	CAA AAA GAA GTC GAA GAA TTT CCA CAC TCA GAA GAA GGG AAT ACT GAG Gln Lys Glu Val Glu Glu Phe Pro His Ser Glu Glu Gly Asn Thr	288
35	TTT CTA GAG ACT ATA GAA GAA AAT AAT TCT GAA GTT CTT GAA CCA GAA Phe Leu Glu Thr Ile Glu Glu Asn Asn Ser Glu Val Leu Glu Pro Glu	336
40	AGG CCT CAA GCA GAA GAA ACC GTT CAG GAA AAA TAT GAG CGC AGT CTT Arg Pro Gln Ala Glu Glu Thr Val Gln Glu Lys Tyr Asp Arg Ser Leu	384
45	AAG AAA ACT CGT ACA GGT TTT GGT GCC CGC TTG AAT GCC TTC TTT GCT Lys Lys Thr Arg Thr Gly Phe Gly Ala Arg Leu Asn Ala Phe Phe Ala	432
50	AAC TTC CGC TCT GTT GAC GAA GAA TTT TTC GAG GAA CTG GAA GAA CTG Asn Phe Arg Ser Val Asp Glu Glu Phe Phe Glu Glu Leu Glu Leu	480
55	CTG ATT ATG AGT GAT GTT GGT GTC CAA GTC GCT TCT AAC TTA ACG GAG Leu Ile Met Ser Asp Val Gly Val Gln Val Ala Ser Asn Leu Thr Glu	528
60	GAA CTA CGT TAC GAA GCC AAG CTT GAA AAT GCC AAG AAA CCT GAT GCA Glu Leu Arg Tyr Glu Ala Lys Leu Glu Asn Ala Lys Lys Pro Asp Ala	576
65	CTT CGT CGT GTC ATC ATT GAG AAA TTG GTT GAG CTT TAT GAA AAG GAT Leu Arg Arg Val Ile Ile Glu Lys Leu Val Glu Leu Thr Glu Lys Asp	624
70	GGT AGC TAC GAT GAA AGC ATC CAC TTC CAA GAT AAC TTG ACA GTT ATG Gly Ser Thr Asp Glu Ser Ile His Phe Gln Asp Asn Leu Thr Val Met	672
75	CTC TTT GTT GGT GTG AAT GGT GTT GGG AAA ACA CACT TCT ATC GGA AAA Leu Phe Val Gly Val Asn Asn Gly Val Gly Lys Thr Thr Ser Ile Gly Lys	720



WO 98/26072

PCT/US97/22578

-206-

	Met	Gly	Leu	Phe	Asp	Arg	Leu	Phe	Gly	Lys	Glu	Glu	Pro	Lys	Ile	
	1				5				10					15		
5	Glu	Glu	Val	Val	Lys	Glu	Ala	Leu	Glu	Asn	Leu	Asp	Leu	Ser	Glu	Asp
				20					25					30		
	Val	Asp	Pro	Thr	Phe	Thr	Glu	Val	Glu	Val	Ser	Gln	Glu	Glu	Ala	
		35					40				45					
10	Glu	Val	Glu	Ile	Val	Glu	Gln	Ala	Val	Phe	Gln	Glu	Glu	Ile	Gln	
	50					55					60					
	Asp	Thr	Val	Glu	Glu	Ser	Leu	Asp	Leu	Glu	Pro	Val	Val	Glu	Val	Ser
	65					70				75				80		
15	Gln	Lys	Glu	Val	Glu	Glu	Phe	Pro	His	Ser	Glu	Glu	Gly	Asn	Thr	Glu
				85					90					95		
	Phe	Leu	Glu	Thr	Ile	Glu	Glu	Asn	Asn	Ser	Glu	Val	Leu	Glu	Pro	Glu
20				100				105						110		
	Arg	Pro	Gln	Ala	Glu	Glu	Thr	Val	Gln	Glu	Lys	Tyr	Asp	Arg	Ser	Leu
		115					120					125				
25	Lys	Lys	Thr	Arg	Thr	Gly	Phe	Gly	Ala	Arg	Leu	Asn	Ala	Phe	Phe	Ala
	130					135					140					
	Asn	Phe	Arg	Ser	Val	Asp	Glu	Glu	Phe	Phe	Glu	Glu	Leu	Glu	Glu	Leu
30					150				155					160		
	Leu	Ile	Met	Ser	Asp	Val	Gly	Val	Gln	Val	Ala	Ser	Asn	Leu	Thr	Glu
				165					170					175		
	Glu	Leu	Arg	Tyr	Glu	Ala	Lys	Leu	Glu	Asn	Ala	Lys	Lys	Pro	Asp	Ala
35				180				185						190		
	Leu	Arg	Arg	Val	Ile	Ile	Glu	Lys	Leu	Val	Glu	Leu	Tyr	Glu	Lys	Asp
		195					200						205			
40	Gly	Ser	Tyr	Asp	Glu	Ser	Ile	His	Phe	Gln	Asp	Asn	Leu	Thr	Val	Met
	210					215						220				
	Leu	Phe	Val	Gly	Val	Asn	Gly	Val	Gly	Lys	Thr	Thr	Ser	Ile	Gly	Lys
45				225		230				235				240		
	Leu	Ala	His	Arg	Tyr	Lys	Gln	Ala	Gly	Lys	Lys	Val	Met	Leu	Val	Ala
				245					250					255		
	Ala	Asp	Thr	Phe	Arg	Ala	Gly	Ala	Val	Ala	Gln	Leu	Ala	Glu	Trp	Gly
50				260			265							270		
	Arg	Arg	Val	Asp	Val	Pro	Val	Val	Thr	Gly	Pro	Glu	Lys	Ala	Asp	Pro
		275					280					285				
55	Ala	Ser	Val	Val	Phe	Asp	Gly	Met	Glu	Arg	Ala	Val	Ala	Glu	Gly	Ile
	290					295					300					
	Asp	Ile	Leu	Met	Ile	Asp	Thr	Ala	Gly	Arg	Leu	Gln	Asn	Lys	Asp	Asn
	305				310					315				320		
60	Leu	Met	Ala	Glu	Leu	Glu	Lys	Ile	Gly	Arg	Ile	Ile	Lys	Arg	Val	Val



WO 98/26072

PCT/US97/22578

-208-

	GAA AAT CAA GAT GGT TTG GAT TTC AAA ATC CTC TTT TTG GAT GCG GCT	288
	Glu Asn Gln Asp Gly Leu Asp Phe Lys Ile Leu Phe Leu Asp Ala Ala	
	85 90 95	
5	GAT AAG GAA TTG GTC GCT CGT TAC AAG GAA ACC AGA CGG AGT CAC CCA	336
	Asp Lys Glu Leu Val Ala Arg Tyr Lys Glu Thr Arg Arg Ser His Pro	
	100 105 110	
10	CTA GCA GCA GAC GGT CGT ATT TTA GAT GGA ATC AAG TTG GAA CGT GAA	384
	Leu Ala Ala Asp Gly Arg Ile Leu Asp Gly Ile Lys Leu Glu Arg Glu	
	115 120 125	
15	CTC TTG GCA CCT TTG AAA AAT ATG AGC CAA AAT GTG GTG GAT ACG ACT	432
	Leu Leu Ala Pro Leu Lys Asn Met Ser Gln Asn Val Val Asp Thr Thr	
	130 135 140	
20	GAA CTC ACT CCA CGT GAG CTG CGC AAA ACC CTT GCA GAG CAG TTT TCA	480
	Glu Leu Thr Pro Arg Glu Leu Arg Lys Thr Leu Ala Glu Gln Phe Ser	
	145 150 155 160	
	GAC CAA GAA CAA GCT CAG TCT TTC CGT ATC GAA GTC ATG TCT TTC GGA	528
	Asp Gln Glu Gln Ala Gln Ser Phe Arg Ile Glu Val Met Ser Phe Gly	
	165 170 175	
25	TTT AAG TAT GGA ATC CCG ATT GAT GCG GAC TTG GTC TTT GAT GTC CGT	576
	Phe Lys Tyr Gly Ile Pro Ile Asp Ala Asp Leu Val Phe Asp Val Arg	
	180 185 190	
30	TTC TTG CCA AAT CCC TAT TAT TTA CCA GAA CTG AGA AAC CAA ACG GGT	624
	Phe Leu Pro Asn Pro Tyr Tyr Leu Pro Glu Leu Arg Asn Gln Thr Gly	
	195 200 205	
35	GTG GAT GAA CCT GTT TAT GAT TAT GTC ATG AAC CAT CCT GAG TCA GAA	672
	Val Asp Glu Pro Val Tyr Asp Tyr Val Met Asn His Pro Glu Ser Glu	
	210 215 220	
40	GAC TTT TAT CAA CAT TTA TTG GCC TTG ATT GAG CCG ATT CTG CCA AGT	720
	Asp Phe Tyr Gln His Leu Leu Ala Leu Ile Glu Pro Ile Leu Pro Ser	
	225 230 235 240	
	TAC CAA AAG GAA GGT AAG TCC GTT TTG ACC ATT GCC ATG GGA TGT ACG	768
	Tyr Gln Lys Glu Gly Lys Ser Val Leu Thr Ile Ala Met Gly Cys Thr	
	245 250 255	
45	GGT GGA CAA CAC CGT AGT GTG GCA TTT GCT AAA CGC TTG GTG CAG GAC	816
	Gly Gly Gln His Arg Ser Val Ala Phe Ala Lys Arg Leu Val Gln Asp	
	260 265 270	
50	TTA TCC AAG AAT TGG TCT GTT AAT GAA GGG CAT CGC GAC AAA GAC CGC	864
	Leu Ser Lys Asn Trp Ser Val Asn Glu Gly His Arg Lys Lys Arg	
	275 280 285	
55	AGA AAG GAA ACG GTA AAC CGT TCA TGA	891
	Arg Lys Glu Thr Val Asn Arg Ser *	
	290 295	

(2) INFORMATION FOR SEQ ID NO:122:

60

(i) SEQUENCE CHARACTERISTICS:

WO 98/26072

PCT/US97/22578

-209-

(A) LENGTH: 296 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

```

10 Met Thr Lys Lys Gln Leu His Leu Val Ile Val Thr Gly Met Gly Gly
    1          5          10          15

    Ala Gly Lys Thr Val Ala Ile Gln Ser Phe Glu Asp Leu Gly Tyr Phe
        20          25          30

15 Thr Ile Asp Asn Met Pro Pro Ala Leu Leu Pro Lys Phe Leu Gln Leu
    35          40          45

    Val Glu Ile Lys Glu Asp Asn Pro Lys Leu Ala Leu Val Val Asp Met
    50          55          60

20 Arg Ser Arg Ser Phe Phe Ser Glu Ile Gln Ala Val Leu Asp Glu Leu
    65          70          75          80

25 Glu Asn Gln Asp Gly Leu Asp Phe Lys Ile Leu Phe Leu Asp Ala Ala
    85          90          95

    Asp Lys Glu Leu Val Ala Arg Tyr Lys Glu Thr Arg Arg Ser His Pro
    100          105          110

30 Leu Ala Ala Asp Gly Arg Ile Leu Asp Gly Ile Lys Leu Glu Arg Glu
    115          120          125

    Leu Leu Ala Pro Leu Lys Asn Met Ser Gln Asn Val Val Asp Thr Thr
    130          135          140

35 Glu Leu Thr Pro Arg Glu Leu Arg Lys Thr Leu Ala Glu Gln Phe Ser
    145          150          155          160

40 Asp Gln Glu Gln Ala Gln Ser Phe Arg Ile Glu Val Met Ser Phe Gly
    165          170          175

    Phe Lys Tyr Gly Ile Pro Ile Asp Ala Asp Leu Val Phe Asp Val Arg
    180          185          190

45 Phe Leu Pro Asn Pro Tyr Tyr Leu Pro Glu Leu Arg Asn Gln Thr Gly
    195          200          205

    Val Asp Glu Pro Val Tyr Asp Tyr Val Met Asn His Pro Glu Ser Glu
    210          215          220

50 Asp Phe Tyr Gln His Leu Leu Ala Leu Ile Glu Pro Ile Leu Pro Ser
    225          230          235          240

55 Tyr Gln Lys Glu Gly Lys Ser Val Leu Thr Ile Ala Met Gly Cys Thr
    245          250          255

    Gly Gly Gln His Arg Ser Val Ala Phe Ala Lys Arg Leu Val Gln Asp
    260          265          270

60 Leu Ser Lys Asn Trp Ser Val Asn Glu Gly His Arg Asp Lys Asp Arg

```

275

280

285

Arg Lys Glu Thr Val Asn Arg Ser  
290 295

5

(2) INFORMATION FOR SEQ ID NO:123:

(1) SEQUENCE CHARACTERISTICS

10

(A) LENGTH: 329 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

15

(11) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

25

Met Val Glu Val Pro Asp Glu Arg Leu Gln Lys Leu Thr Glu Met Ile  
1 5 10 15

Thr Pro Lys Lys Thr Val Pro Thr Thr Phe Glu Phe Thr Asp Ile Ala  
20 25 30

30

Gly Ile Val Lys Gly Ala Ser Lys Gly Glu Gly Leu Gly Asn Lys Phe  
35 40 45

35

Leu Ala Asn Ile Arg Glu Val Asp Ala Ile Val His Val Val Arg Ala  
50 55 60

Phe Asp Asp Glu Asn Val Met Arg Glu Gln Gly Arg Glu Asp Ala Phe  
65 70 75 80

40

Val Asp Pro Leu Ala Asp Ile Asp Thr Ile Asn Leu Glu Leu Ile Leu  
85 90 95

Ala Asp Leu Glu Ser Val Asn Lys Arg Tyr Ala Arg Val Glu Lys Met  
100 105 110

45

Ala Arg Thr Gln Lys Asp Lys Glu Ser Val Ala Glu Phe Asn Val Leu  
115 120 125

Gln Lys Ile Lys Pro Val Leu Glu Asp Gly Lys Ser Ala Arg Thr Ile  
130 135 140

50

Glu Phe Thr Asp Glu Glu Gln Lys Val Val Lys Gly Leu Phe Leu Leu  
145 150 155 160

59

Thr Thr Lys Pro Val Leu Tyr Val Ala Asn Val Asp Glu Asp Val Val  
165 170 175

Ser Glu Pro Asp Ser Ile Asp Tyr Val Lys Gln Ile Arg Glu Phe Ala  
180 185 190

60

Ala Thr Glu Asp Ala Glu Val Val Val Ile Ser Ala Arg Ala Glu Glu



-211-

	195	200	205
	Glu Ile Ser Glu Leu Asp	Glu Asp Lys Lys	Glu Phe Leu Glu Ala
	210	215	220
5	Ile Gly Leu Thr Glu Ser Gly Val Asp Lys Leu Thr Arg Ala Ala Tyr		
	225	230	235
10	His Leu Leu Gly Leu Gly Thr Tyr Phe Thr Ala Gly Glu Lys Glu Val		
	245	250	255
	Arg Ala Trp Thr Phe Lys Arg Gly Met Lys Ala Pro Gln Ala Ala Gly		
	260	265	270
15	Ile Ile His Ser Asp Phe Glu Lys Gly Phe Ile Arg Ala Val Thr Met		
	275	280	285
	Ser Tyr Glu Asp Leu Val Lys Tyr Gly Ser Glu Lys Ala Val Lys Glu		
	290	295	300
20	Ala Gly Arg Leu Arg Glu Glu Gly Lys Glu Tyr Ile Val Gln Asp Gly		
	305	310	315
25	Asp Ile Met Glu Phe Arg Phe Asn Val		
	325		

(2) INFORMATION FOR SEQ ID NO:124:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 189 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

Met Ser Ala Ser Glu Gly Arg Asp Pro Tyr Glu Asp Tyr Leu Ala Ile	
1	5
Asn Lys Glu Leu Glu Ser Tyr Asn Leu Arg Leu Met Glu Arg Pro Gln	
20	25
Ile Ile Val Thr Asn Lys Met Asp Met Pro Glu Ser Gln Glu Asn Leu	
35	40
Glu Glu Phe Lys Lys Lys Leu Ala Glu Asn Tyr Asp Glu Phe Glu Glu	
50	55
Leu Pro Ala Ile Phe Pro Ile Ser Gly Leu Thr Lys Gln Gly Leu Ala	
65	70
Thr Leu Leu Asp Ala Thr Ala Glu Leu Leu Asp Lys Thr Pro Glu Phe	
85	90

WO 98/26072

PCT/US97/22578

-212-

Leu Leu Tyr Asp Glu Ser Asp Met Glu Glu Val Tyr Tyr Gly Phe  
100 105 110

5 Asp Glu Glu Glu Lys Ala Phe Glu Ile Ser Arg Asp Asp Ala Thr  
115 120 125

Trp Val Leu Ser Gly Glu Lys Leu Met Lys Leu Phe Asn Met Thr Asn  
130 135 140

10 Phe Asp Arg Asp Glu Ser Val Met Lys Phe Ala Arg Gln Leu Arg Gly  
145 150 155 160

Met Gly Val Asp Glu Ala Leu Arg Ala Arg Gly Ala Lys Asp Gly Asp  
165 170 175

15 Leu Val Arg Ile Gly Lys Phe Glu Phe Glu Phe Val Asp  
180 185

(2) INFORMATION FOR SEQ ID NO:125:

20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 226 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
25 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

30 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

35 Met Asn Ile Gln Gln Leu Arg Tyr Val Val Ala Ile Ala Asn Ser Gly  
1 5 10 15

40 Thr Phe Arg Glu Ala Ala Glu Lys Met Tyr Val Ser Gln Pro Ser Leu  
20 25 30

Ser Ile Ser Val Arg Asp Leu Glu Lys Glu Leu Gly Phe Lys Ile Phe  
35 40 45

45 Arg Arg Thr Ser Ser Gly Thr Phe Leu Thr Arg Arg Gly Met Glu Phe  
50 55 60

Tyr Glu Lys Ala Gln Glu Leu Val Lys Gly Phe Asp Ile Phe Gln Asn  
65 70 75 80

50 Gln Tyr Ala Asn Pro Glu Glu Glu Lys Asp Glu Phe Ser Val Ala Ser  
85 90 95

Gln His Tyr Asp Phe Leu Pro Pro Thr Ile Thr Ala Phe Ser Glu Arg  
100 105 110

55 Tyr Pro Asp Tyr Lys Asn Phe Arg Ile Phe Glu Ser Thr Thr Val Gln  
115 120 125

60 Ile Leu Asp Glu Val Ala Gln Gly His Ser Glu Ile Gly Ile Ile Tyr  
130 135 140

WO 98/26072

PCT/US97/22578

-213-

5 Leu Asn Asn Gln Asn Lys Lys Gly Ile Met Gln Arg Val Glu Lys Leu  
145 150 155 160

10 Gly Leu Glu Val Ile Glu Leu Ile Pro Phe His Thr His Ile Tyr Leu  
165 170 175

15 Cys Glu Gly His Pro Leu Ala Gln Lys Glu Glu Leu Val Met Glu Asp  
180 185 190

20 Leu Ala Asp Leu Pro Thr Val Arg Phe Thr Gln Lys Asp Glu Tyr  
195 200 205

25 Leu Tyr Tyr Ser Glu Asn Phe Val Asp Thr Ser Ala Thr His Arg Cys  
210 215 220

30 Leu Met  
225

(2) INFORMATION FOR SEQ ID NO:126:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 119 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

Met Lys Lys Arg Ala Ile Val Ala Val Ile Val Leu Leu Leu Ile Gly  
1 5 10 15

40 Leu Asp Gln Leu Val Lys Ser Tyr Ile Val Gln Gln Ile Pro Leu Gly  
20 25 30

Glu Val Arg Ser Trp Ile Pro Asn Phe Val Ser Leu Thr Tyr Leu Gln  
35 40 45

45 Asn Arg Gly Ala Ala Phe Ser Ile Leu Gln Asp Gln Gln Leu Leu Phe  
50 55 60

50 Ala Val Ile Thr Leu Val Val Ile Gly Ala Ile Trp Tyr Leu His  
65 70 75 80

Lys His Met Glu Asp Ser Phe Trp Met Val Leu Gly Leu Thr Leu Ile  
85 90 95

55 Ile Ala Gly Gly Pro Gly Asn Phe Ile Asp Arg Val Ser Gln Gly Phe  
100 105 110

60 Val Val Asp Met Phe His Leu  
115

(2) INFORMATION FOR SEQ ID NO:127:

-214-

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 170 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:
- |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ser | Lys | Tyr | Leu | Leu | Lys | Leu | Leu | Val | Tyr | Cys | Phe | Ser | Ala | Leu | 1   | 5   | 10  | 15  |
| Thr | Phe | Gly | Ser | Leu | Phe | Leu | Ile | Ile | Gly | Phe | Ile | Leu | Ile | Lys | Gly | 20  | 25  | 30  |     |
| Leu | Pro | His | Leu | Ser | Leu | Ser | Leu | Phe | Ser | Trp | Thr | Tyr | Thr | Ser | Glu | 35  | 40  | 45  |     |
| Asn | Ile | Ser | Leu | Met | Pro | Ala | Ile | Ile | Ser | Thr | Val | Ile | Leu | Val | Phe | 50  | 55  | 60  |     |
| Gly | Ala | Leu | Leu | Leu | Ala | Leu | Pro | Ile | Gly | Ile | Phe | Ala | Gly | Phe | Tyr | 65  | 70  | 75  | 80  |
| Leu | Val | Glu | Tyr | Thr | Lys | Lys | Asp | Ser | Leu | Cys | Val | Lys | Ile | Met | Arg | 85  | 90  | 95  |     |
| Leu | Ala | Ser | Asp | Thr | Leu | Ser | Gly | Ile | Pro | Ser | Ile | Val | Phe | Gly | Leu | 100 | 105 | 110 |     |
| Phe | Gly | Met | Leu | Phe | Phe | Val | Val | Phe | Leu | Gly | Phe | Gln | Tyr | Ser | Leu | 115 | 120 | 125 |     |
| Leu | Ser | Gly | Ile | Leu | Thr | Ser | Val | Ile | Met | Val | Leu | Pro | Val | Ile | Ile | 130 | 135 | 140 |     |
| Arg | Ser | Thr | Glu | Glu | Ala | Leu | Leu | Ser | Val | Ser | Asp | Ser | Met | Arg | Gln | 145 | 150 | 155 | 160 |
| Ala | Ser | Tyr | Gly | Leu | Gly | Ala | Leu | Ser | Tyr |     |     |     |     |     |     | 165 | 170 |     |     |
- (2) INFORMATION FOR SEQ ID NO:128:
- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 154 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO

-215-

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

5 Met Lys Thr Glu Gln Thr Ala Ser Lys Thr Ser Ala Leu Lys Gly Lys  
1 5 10 15

10 Glu Val Ala Asp Phe Glu Leu Met Gly Val Asp Gly Lys Thr Tyr Arg  
20 25 30

Leu Ser Asp Tyr Lys Gly Lys Lys Val Tyr Leu Lys Phe Trp Ala Ser  
35 40 45

15 Trp Cys Ser Ile Cys Leu Ala Ser Leu Pro Asp Thr Asp Glu Ile Ala  
50 55 60

Lys Glu Ala Gly Asp Asp Tyr Val Val Leu Thr Val Val Ser Pro Gly  
65 70 75 80

20 His Lys Gly Glu Gln Ser Glu Ala Asp Phe Lys Asn Trp Tyr Lys Gly  
85 90 95

25 Leu Asp Tyr Lys Asn Leu Pro Val Leu Val Asp Pro Ser Gly Lys Leu  
100 105 110

Leu Glu Thr Tyr Gly Val Arg Ser Tyr Pro Thr Gln Ala Phe Ile Asp  
115 120 125

30 Lys Glu Gly Lys Leu Val Lys Thr His Pro Gly Phe Met Glu Lys Asp  
130 135 140

Ala Ile Leu Gln Thr Leu Lys Glu Leu Ser  
145 150

(2) INFORMATION FOR SEQ ID NO:129:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 181 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

55 Met Lys Lys Glu Gln Ile Pro Asn Leu Leu Thr Ile Gly Arg Ile Leu  
1 5 10 15

Phe Ile Pro Ile Phe Ile Phe Ile Leu Thr Ile Gly Asn Ser Ile Glu  
20 25 30

Ser His Ile Val Ala Ala Ile Ile Phe Ala Val Ala Ser Ile Thr Asp

WO 98/26072

PCT/US97/22578

-216-

	35	40	45
5	Tyr Leu Asp Gly Tyr Leu Ala Arg Lys Trp Asn Val Val Ser Asn Phe 50 55 60		
	Gly Lys Phe Ala Asp Pro Met Ala Asp Lys Leu Leu Val Met Ser Ala 65 70 75 80		
10	Phe Ile Met Leu Ile Glu Leu Gly Met Ala Pro Ala Trp Ile Val Ala 85 90 95		
	Val Ile Ile Cys Arg Glu Leu Ala Val Thr Gly Leu Arg Leu Leu Leu 100 105 110		
15	Val Glu Thr Gly Gly Thr Ile Leu Ala Ala Ala Met Pro Gly Lys Ile 115 120 125		
	Lys Thr Phe Ser Gln Met Phe Ala Ile Ile Phe Leu Leu Leu His Trp 130 135 140		
20	Thr Leu Leu Gly Gln Val Leu Leu Tyr Val Ala Leu Phe Phe Thr Ile 145 150 155 160		
25	Tyr Ser Gly Tyr Asp Tyr Phe Lys Gly Ser Ala Tyr Val Phe Lys Gly 165 170 175		
	Thr Phe Gly Ser Lys 180		

30 (2) INFORMATION FOR SEQ ID NO:130:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 111 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

35 (ii) MOLECULE TYPE: peptide

40 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

50	Leu Arg Leu Lys Glu Met Asn Gly Asp Met Ile His Ala Ala Tyr Asp 1 5 10 15
	Leu Gly Ala Ser Gln Phe Gln Met Phe Lys Glu Ile Met Leu Pro Tyr 20 25 30
55	Leu Thr Pro Ser Ile Ile Ala Gly Tyr Phe Met Ala Phe Thr Tyr Ser 35 40 45
	Leu Asp Asp Phe Ala Val Thr Phe Phe Val Thr Gly Asn Gly Phe Ser 50 55 60
60	Thr Leu Ser Val Glu Ile Tyr Ser Arg Ala Arg Lys Gly Ile Ser Leu

WO 98/26072

PCT/US97/22578

-217-

```

65          70          75          80
Glu Ile Asn Ala Leu Ser Ala Leu Val Phe Leu Phe Ser Ile Ile Leu
      85          90          95
5 Val Val Gly Tyr Tyr Phe Ile Ser Arg Glu Lys Glu Glu Gln Ala
      100      105      110

(2) INFORMATION FOR SEQ ID NO:131:

10 (i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 59 amino acids
    (B) TYPE: amino acid
    (C) STRANDEDNESS: not relevant
    (D) TOPOLOGY: not relevant

15 (ii) MOLECULE TYPE: peptide

    (iii) HYPOTHETICAL: NO

20 (iv) ANTI-SENSE: NO

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

25 Pro Gln Phe Thr Glu Glu Thr Gly Ile Gln Val Gln Tyr Glu Ala Phe
    1          5          10          15

30 Asp Ser Asn Glu Ala Met Tyr Thr Lys Ile Lys Gln Gly Glu Thr Thr
    20      25      30

Tyr Asp Ile Ala Ile Pro Ser Glu Tyr Met Ile Asn Lys Met Lys Asp
    35      40      45

35 Glu Asp Leu Leu Val Pro Leu Asp Tyr Ser Lys
    50      55

(2) INFORMATION FOR SEQ ID NO:132:

40 (i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 232 amino acids
    (B) TYPE: amino acid
    (C) STRANDEDNESS: not relevant
    (D) TOPOLOGY: not relevant

45 (ii) MOLECULE TYPE: peptide

    (iii) HYPOTHETICAL: NO

50 (iv) ANTI-SENSE: NO

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

55 Met Gln Thr Gln Glu Lys His Ser Gln Ala Ala Val Leu Gly Leu Gln
    1          5          10          15

His Leu Leu Ala Met Tyr Ser Gly Ser Ile Leu Val Pro Ile Met Ile
    20      25      30

60 Ala Thr Ala Leu Gly Tyr Ser Ala Glu Gln Leu Thr Tyr Leu Ile Ser

```





PCT/US97/22578

[illegible]

(2) INFORMATION FOR SEQ ID NO:134:

-220-

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 184 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

Ser	Leu	Ile	Ile	Ala	Leu	Ala	Thr	Thr	Leu	Ile	Ala	Ile	Ile	Ile	Ser
1				5					10					15	
Ala	Met	Ala	Ala	Tyr	Gly	Ile	Val	Arg	Phe	Phe	Pro	Lys	Leu	Gly	Ala
				20				25					30		
Ile	Met	Ser	Arg	Leu	Leu	Val	Ile	Thr	Tyr	Ile	Phe	Pro	Pro	Ile	Leu
			35			40						45			
Leu	Ala	Ile	Pro	Tyr	Ser	Ile	Ala	Ile	Ala	Lys	Val	Gly	Leu	Thr	Asn
		50				55				60					
Ser	Leu	Phe	Gly	Leu	Met	Met	Val	Tyr	Leu	Ser	Phe	Ser	Val	Pro	Tyr
65				70					75				80		
Ala	Val	Trp	Leu	Leu	Val	Gly	Phe	Phe	Gln	Thr	Val	Pro	Ile	Gly	Ile
			85					90					95		
Glu	Glu	Ala	Ala	Arg	Ile	Asp	Gly	Ala	Asn	Lys	Phe	Val	Thr	Phe	Tyr
			100				105						110		
Lys	Val	Val	Leu	Pro	Ile	Val	Ala	Pro	Gly	Ile	Val	Ala	Thr	Ala	Ile
		115				120					125				
Tyr	Thr	Phe	Ile	Asn	Ala	Trp	Asn	Glu	Phe	Leu	Tyr	Ala	Leu	Ile	Leu
		130				135					140				
Ile	Asn	Asn	Thr	Gly	Lys	Met	Thr	Val	Ala	Val	Ala	Leu	Arg	Ser	Leu
145				150					155				160		
Asn	Gly	Ser	Glu	Ile	Leu	Asp	Trp	Gly	Asp	Met	Met	Ala	Ala	Ser	Val
			165					170					175		
Ile	Val	Val	Leu	Pro	Ser	Ile	Ile								
			180												

(2) INFORMATION FOR SEQ ID NO:135:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 155 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

WO 98/26072

PCT/US97/22578

-221-

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

Asp Glu Leu Ala Asp Leu Met Met Val Ala Ser Lys Glu Val Glu Asp  
1 5 10 15  
Ala Ile Ile Arg Leu Gly Gln Lys Ala Arg Ala Ala Gly Ile His Met  
20 25 30  
Ile Leu Ala Thr Gln Arg Pro Ser Val Asp Val Ile Ser Gly Leu Ile  
35 40 45  
Lys Ala Asn Val Pro Ser Arg Val Ala Phe Ala Val Ser Ser Gly Thr  
50 55 60  
Asp Ser Arg Thr Ile Leu Asp Glu Asn Gly Ala Glu Lys Leu Leu Gly  
65 70 75 80  
Arg Gly Asp Met Leu Phe Lys Pro Ile Asp Glu Asn His Pro Val Arg  
85 90 95  
Leu Gln Gly Ser Phe Ile Ser Asp Asp Asp Val Glu Arg Ile Val Asn  
100 105 110  
Phe Ile Lys Thr Gln Ala Asp Ala Asp Tyr Asp Glu Ser Phe Asp Pro  
115 120 125  
Gly Glu Val Ser Glu Asn Glu Gly Glu Phe Ser Asp Gly Asp Ala Gly  
130 135 140  
Gly Asp Pro Leu Phe Glu Glu Ala Lys Ser Leu  
145 150 155

(2) INFORMATION FOR SEQ ID NO:136:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 154 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

Met Thr Glu Asn Thr Pro Lys Ala Leu Val Gln Val Asn Gln Lys Pro  
1 5 10 15  
Leu Ile Glu Tyr Gln Ile Glu Phe Leu Lys Glu Lys Gly Ile Asn Asp  
20 25 30

WO 98/26072

PCT/US97/22578

-222-

Ile Ile Ile Ile Val Gly Tyr Leu Lys Glu Gln Phe Asp Tyr Leu Lys  
35 40 45  
5 Glu Lys Tyr Gly Val Arg Leu Val Phe Asn Asp Lys Tyr Ala Asp Tyr  
50 55 60  
Asn Asn Phe Tyr Ser Leu Tyr Leu Val Lys Glu Leu Ala Asn Ser  
65 70 75 80  
10 Tyr Val Ile Asp Ala Asp Asn Tyr Leu Phe Lys Asn Met Phe Arg Asn  
85 90 95  
Asp Leu Thr Arg Ser Thr Tyr Phe Ser Val Tyr Arg Glu Asp Cys Thr  
100 105 110  
Asn Glu Trp Phe Leu Val Tyr Gly Asp Asp Tyr Lys Val Gln Asp Ile  
115 120 125  
20 Ile Val Asp Ser Lys Ala Gly Arg Ile Leu Ser Gly Val Ser Phe Trp  
130 135 140  
Asp Ala Pro Thr Ala Glu Lys Ile Val Ser  
145 150  
25

(2) INFORMATION FOR SEQ ID NO:137:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 286 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant  
30  
(ii) MOLECULE TYPE: peptide  
35  
(iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: NO  
40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

Met Ser Asp Asn Ser Lys Thr Arg Val Val Gly Met Ser Gly Gly  
1 5 10 15  
Val Asp Ser Ser Val Thr Ala Leu Leu Leu Lys Glu Gln Gly Tyr Asp  
20 25 30  
50 Val Ile Gly Ile Phe Met Lys Asn Trp Asp Asp Thr Asp Glu Asn Gly  
35 40 45  
Val Cys Thr Ala Thr Glu Asp Tyr Lys Asp Val Val Ala Val Ala Asp  
50 55 60  
Gln Ile Gly Ile Pro Tyr Tyr Ser Val Asn Phe Glu Lys Glu Tyr Trp  
65 70 75 80  
60 Asp Arg Val Phe Glu Tyr Phe Leu Ala Glu Tyr Arg Ala Gly Arg Thr  
85 90 95

WO 98/26072

PCT/US97/22578

-223-

	Pro	Asn	Pro	Asp	Val	Met	Cys	Asn	Lys	Glu	Ile	Lys	Phe	Lys	Ala	Phe	
				100					105						110		
5	Leu	Asp	Tyr	Ala	Met	Thr	Leu	Gly	Ala	Asp	Tyr	Val	Ala	Thr	Gly	His	
				115				120					125				
	Tyr	Ala	Arg	Val	Ala	Arg	Asp	Glu	Asp	Gly	Thr	Val	His	Met	Leu	Arg	
				130				135					140				
10	Gly	Val	Asp	Asn	Gly	Lys	Asp	Gln	Thr	Tyr	Phe	Leu	Ser	Gln	Leu	Ser	
						150						155				160	
	Gln	Glu	Gln	Leu	Gln	Lys	Thr	Met	Phe	Pro	Leu	Gly	His	Leu	Lys	Lys	
						165				170					175		
15	Pro	Glu	Val	Arg	Lys	Leu	Ala	Glu	Glu	Ala	Gly	Leu	Ser	Thr	Ala	Lys	
						180				185					190		
20	Lys	Lys	Asp	Ser	Thr	Gly	Ile	Cys	Phe	Ile	Gly	Glu	Lys	Asn	Phe	Lys	
						195			200				205				
	Asn	Phe	Leu	Ser	Asn	Tyr	Leu	Pro	Ala	Gln	Pro	Gly	Arg	Met	Met	Thr	
							210		215				220				
25	Val	Asp	Gly	Arg	Asp	Met	Gly	Glu	His	Ala	Gly	Leu	Met	Tyr	Tyr	Thr	
							230				235					240	
	Ile	Gly	Gln	Arg	Gly	Gly	Leu	Gly	Ile	Gly	Gly	Gln	His	Gly	Gly	Asp	
						245				250					255		
30	Asn	Ala	Pro	Trp	Phe	Val	Val	Gly	Lys	Asp	Leu	Ser	Lys	Asn	Ile	Leu	
							260			265					270		
35	Tyr	Val	Gly	Gln	Gly	Phe	Tyr	His	Asp	Ser	Leu	Met	Ser	Thr			
								280						285			

(2) INFORMATION FOR SEO ID NO:138:

40 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 648 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

45 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

50 (iv) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:138:

55 Met Glu Val Phe Glu Ser Leu Lys Ala Asn Leu Val Gly Lys Asn Ala  
1 5 10 15

Arg Ile Val Leu Pro Glu Gly Glu Glu Pro Arg Ile Leu Gln Ala Thr  
20 25 30

60 Lys Arg Leu Val Lys Glu Thr Glu Val Ile Pro Val Leu Leu Gly Asn

WO 98/26072

PCT/US97/22578

-224-

	35	40	45
5	Pro Glu Lys Ile Lys Ile Tyr 50	Leu Glu Ile Glu Gly Ile Met Asp Gly 55	
	Tyr Glu Val Ile Asp Pro Gln His Tyr Pro Gln Phe Glu Glu Met Val 65	70	75 80
10	Ser Ala Leu Val Glu Arg Arg Lys Gly Lys Met Thr Glu Glu Asp Val 85	90	95
	Arg Lys Val Leu Val Glu Asp Val Asn Tyr Phe Gly Val Met Leu Val 100	105	110
15	Tyr Leu Gly Leu Val Asp Gly Met Val Ser Gly Ala Ile His Ser Thr 115	120	125
	Ala Ser Thr Val Arg Pro Ala Leu Gln Ile Ile Lys Thr Arg Pro Asn 130	135	140
20	Val Thr Arg Thr Ser Gly Ala Phe Leu Met Val Arg Gly Thr Glu Arg 145	150	155
	Tyr Leu Phe Gly Asp Cys Ala Ile Asn Ile Asn Pro Asp Ala Glu Ala 165	170	175
25	Leu Ala Glu Ile Ala Ile Asn Ser Ala Ile Thr Ala Lys Met Phe Gly 180	185	190
30	Ile Glu Pro Lys Ile Ala Met Leu Ser Tyr Ser Thr Lys Gly Ser Gly 195	200	205
	Phe Gly Glu Ser Val Asp Lys Val Val Glu Ala Thr Lys Ile Ala His 210	215	220
35	Asp Leu Arg Pro Asp Leu Glu Ile Asp Gly Glu Leu Gln Phe Asp Ala 225	230	235
	Ala Phe Val Pro Glu Thr Ala Ala Leu Lys Ala Pro Gly Ser Thr Val 245	250	255
40	Ala Gly Gln Ala Asn Val Phe Ile Phe Pro Gly Ile Glu Ala Gly Asn 260	265	270
45	Ile Gly Tyr Lys Met Ala Glu Arg Leu Gly Gly Phe Ala Ala Val Gly 275	280	285
	Pro Val Leu Gln Gly Leu Asn Lys Pro Val Asn Asp Leu Ser Arg Gly 290	295	300
50	Cys Asn Ala Asp Asp Val Tyr Lys Leu Thr Leu Ile Thr Ala Ala Gln 305	310	315
	Ala Val His Gln Met Glu Val Phe Glu Ser Leu Lys Ala Asn Leu Val 325	330	335
55	Gly Lys Asn Ala Arg Ile Val Leu Pro Glu Gly Glu Glu Pro Arg Ile 340	345	350
60	Leu Gln Ala Thr Lys Arg Leu Val Lys Glu Thr Glu Val Ile Pro Val 355	360	365

WO 98/26072

PCT/US97/22578

-225-

Leu Leu Gly Asn Pro Glu Lys Ile Lys Ile Tyr Leu Glu Ile Glu Gly  
 370 375 380  
 5 Ile Met Asp Gly Tyr Glu Val Ile Asp Pro Gln His Tyr Pro Gln Phe  
 385 390 395 400  
 Glu Glu Met Val Ser Ala Leu Val Glu Arg Arg Lys Gly Lys Met Thr  
 405 410 415  
 10 Glu Glu Asp Val Arg Lys Val Leu Val Glu Asp Val Asn Tyr Phe Gly  
 420 425 430  
 Val Met Leu Val Tyr Leu Gly Leu Val Asp Gly Met Val Ser Gly Ala  
 435 440 445  
 15 Ile His Ser Thr Ala Ser Thr Val Arg Pro Ala Leu Gln Ile Ile Lys  
 450 455 460  
 20 Thr Arg Pro Asn Val Thr Arg Thr Ser Gly Ala Phe Leu Met Val Arg  
 465 470 475 480  
 Gly Thr Glu Arg Tyr Leu Phe Gly Asp Cys Ala Ile Asn Ile Asn Pro  
 485 490 495  
 25 Asp Ala Glu Ala Leu Ala Glu Ile Ala Ile Asn Ser Ala Ile Thr Ala  
 500 505 510  
 Lys Met Phe Gly Ile Glu Pro Lys Ile Ala Met Leu Ser Tyr Ser Thr  
 515 520 525  
 30 Lys Gly Ser Gly Phe Gly Glu Ser Val Asp Lys Val Val Glu Ala Thr  
 530 535 540  
 35 Lys Ile Ala His Asp Leu Arg Pro Asp Leu Glu Ile Asp Gly Glu Leu  
 545 550 555 560  
 Gln Phe Asp Ala Ala Phe Val Pro Glu Thr Ala Ala Leu Lys Ala Pro  
 565 570 575  
 40 Gly Ser Thr Val Ala Gly Gln Ala Asn Val Phe Ile Phe Pro Gly Ile  
 580 585 590  
 Glu Ala Gly Asn Ile Gly Tyr Lys Met Ala Glu Arg Leu Gly Gly Phe  
 595 600 605  
 45 Ala Ala Val Gly Pro Val Leu Gln Gly Leu Asn Lys Pro Val Asn Asp  
 610 615 620  
 50 Leu Ser Arg Gly Cys Asn Ala Asp Asp Val Tyr Lys Leu Thr Leu Ile  
 625 630 635 640  
 Thr Ala Ala Gln Ala Val His Gln  
 645  
 55

(2) INFORMATION FOR SEQ ID NO:139:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 135 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: not relevant

60

WO 98/26072

PCT/US97/22578

-226-

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

5 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

Met	Arg	Asn	Leu	Lys	Ser	Ile	Leu	Arg	Arg	His	Ile	Ser	Leu	Leu	Gly
1			5					10						15	

Phe	Leu	Gly	Val	Leu	Ser	Ile	Trp	Gln	Leu	Ala	Gly	Phe	Leu	Lys	Leu
			20				25				30				

Leu	Pro	Lys	Phe	Ile	Leu	Pro	Thr	Pro	Leu	Glu	Ile	Leu	Gln	Pro	Phe
		35				40					45				

Val	Arg	Asp	Arg	Glu	Phe	Leu	Trp	His	His	Ser	Trp	Ala	Thr	Leu	Arg
	50				55						60				

Val	Ala	Leu	Leu	Gly	Leu	Ile	Leu	Gly	Val	Leu	Ile	Ala	Cys	Leu	Met
	65			70				75						80	

Ala	Val	Leu	Met	Asp	Ser	Leu	Thr	Trp	Leu	Asn	Asp	Leu	Ile	Tyr	Pro
			85					90					95		

Met	Met	Val	Val	Ile	Gln	Thr	Ile	Pro	Thr	Ile	Ala	Ile	Ala	Pro	Ile
		100					105						110		

Leu	Val	Leu	Trp	Leu	Gly	Tyr	Gly	Ile	Phe	Ala	Gln	Asp	Cys	Leu	Asp
		115			120							125			

Tyr	Leu	Asn	Asn	Asn	Leu	Ser
	130				135	

(2) INFORMATION FOR SEQ ID NO:140:

40

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 279 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: not relevant

45

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

50

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

55

Pro	Trp	Ser	Leu	Val	Asp	Glu	Tyr	Glu	Gln	Leu	Tyr	Ala	Thr	Ile	Gly
	1			5				10						15	

Trp	His	Pro	Thr	Glu	Ala	Gly	Thr	Tyr	Thr	Glu	Glu	Val	Glu	Ala	Tyr
		20				25							30		

60



WO 98/26072

PCT/US97/22578

-227-

Leu Leu Asp Lys Leu Lys His Ser Lys Val Val Ala Leu Gly Glu Ile  
 35 40 45  
 Gly Leu Asp Tyr His Trp Met Thr Ala Pro Glu Val Gln Glu Gln Val  
 50 55 60  
 Phe Arg Arg Gln Ile Gln Leu Ser Lys Asp Leu Asp Leu Pro Phe Val  
 65 70 75 80  
 Val His Thr Arg Asp Ala Leu Glu Asp Thr Tyr Glu Ile Ile Lys Ser  
 85 90 95  
 Glu Gly Val Gly Pro Arg Gly Gly Ile Met His Ser Phe Ser Gly Thr  
 100 105 110  
 Leu Glu Trp Ala Arg Tyr Arg Asp Leu Gly Met Thr Ile Ser Phe Ser  
 115 120 125  
 Gly Val Val Thr Phe Lys Lys Ala Thr Asp Leu Gln Glu Ala Ala Lys  
 130 135 140  
 Glu Leu Pro Leu Asp Lys Met Leu Val Glu Thr Asp Ala Pro Tyr Leu  
 145 150 155 160  
 Ala Pro Val Pro Lys Arg Gly Arg Glu Asn Lys Thr Ala Tyr Thr Arg  
 165 170 175  
 Tyr Val Val Asp Phe Ile Ala Asp Leu Arg Gly Met Thr Thr Glu Glu  
 180 185 190  
 Leu Ala Val Ala Thr Thr Ala Asn Ala Glu Arg Ile Phe Gly Ile Gly  
 195 200 205  
 Gln Gln Val Met Lys Glu Arg Ile Ser Gln Val Ile Val Val Glu Gly  
 210 215 220  
 Arg Asp Asp Thr Val Asn Leu Lys Arg Tyr Phe Asp Val Glu Thr Tyr  
 225 230 235 240  
 Glu Thr Arg Gly Ser Ala Ile Asn Asp Gln Asp Ile Glu Arg Ile Gln  
 245 250 255  
 Arg Leu His Gln Arg His Gly Val Ile Val Phe Thr Asp Pro Asp Phe  
 260 265 270  
 Asn Gly Asp Gly Phe Gly Ala  
 275

- (2) INFORMATION FOR SEQ ID NO:141:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 147 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

WO 98/26072

PCT/US97/22578

-228-

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:141:

5 Met Lys Ile Ile Ile Gln Arg Val Lys Lys Ala Gln Val Ser Ile Glu  
1 5 10  
Gly Gln Ile Gln Gly Lys Ile Asn Gln Gly Leu Leu Leu Val Gly  
20 25  
10 Val Gly Pro Glu Asp Gln Glu Glu Asp Leu Asp Tyr Ala Val Arg Lys  
35 40 45  
15 Leu Val Asn Met Arg Ile Phe Ser Asp Ala Glu Gly Lys Met Asn Leu  
50 55 60  
Ser Val Lys Asp Ile Glu Gly Glu Ile Leu Ser Ile Ser Gln Phe Thr  
65 70 75 80  
20 Leu Phe Ala Asp Thr Lys Lys Gly Asn Arg Pro Ala Phe Thr Gly Ala  
85 90 95  
Ala Lys Pro Asp Met Ala Ser Asp Phe Tyr Asp Ala Phe Asn Gln Lys  
100 105 110  
25 Leu Ala Gln Glu Val Pro Val Gln Thr Gly Ile Phe Gly Ala Asp Met  
115 120 125  
Gln Val Glu Leu Val Asn Asn Gly Pro Val Thr Ile Ile Leu Asp Thr  
130 135 140  
Lys Lys Arg  
145

35 (2) INFORMATION FOR SEQ ID NO:142:

(i) SEQUENCE CHARACTERISTICS:

40 (A) LENGTH: 238 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

45 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

Met Ile Leu Ser Met Val Ser Thr Pro Leu Pro Ser Ser Pro Cys Lys  
1 5 10  
55 Tyr Arg Lys Gln Leu Tyr Leu Gln Glu Asp Leu Arg Gly Lys Asn Val  
20 25  
Glu Lys Val Lys Glu Leu Ala Thr Glu Lys Lys Val Ser Ile Ser Trp  
35 40 45  
60 Thr Ser Lys Lys Ser Leu Ser Glu Met Thr Glu Gly Ala Val His Gln

WO 98/26072

PCT/US97/22578

-229-

	50		55		60
	Gly Phe Val Leu Arg Val Ser Glu Phe Ala Tyr Ser Glu Leu Asp Tyr				
	65		70		75
5	Ile Leu Ala Lys Thr Arg Gln Glu Glu Asn Pro Leu Leu Leu Ile Leu				
		85		90	95
10	Asp Gly Leu Thr Asp Pro His Asn Leu Gly Ser Ile Leu Arg Thr Ala				
		100		105	110
	Asp Ala Thr Asn Val Ser Gly Val Ile Ile Pro Lys His Arg Ala Val				
		115		120	125
15	Gly Val Thr Pro Val Val Ala Lys Thr Ala Thr Gly Ala Ile Glu His				
		130		135	140
	Val Pro Ile Ala Arg Val Thr Asn Leu Ser Gln Thr Leu Asp Lys Leu				
		145		150	155
20	Lys Asp Glu Gly Phe Thr Thr Phe Gly Thr Asp Met Asn Gly Thr Pro				
		165		170	175
25	Cys His Lys Trp Asn Thr Lys Gly Lys Ile Ala Leu Ile Ile Gly Asn				
		180		185	190
	Glu Gly Lys Gly Ile Ser Ser Asn Ile Lys Lys Gln Val Asp Glu Met				
		195		200	205
30	Ile Thr Ile Pro Met Asn Gly His Val Gln Ser Leu Asn Ala Ser Val				
		210		215	220
	Ala Ala Ala Ile Leu Met Tyr Glu Val Phe Arg Asn Arg Leu				
		225		230	235
35	(2) INFORMATION FOR SEQ ID NO:143:				
	(i) SEQUENCE CHARACTERISTICS:				
	(A) LENGTH: 345 amino acids				
40	(B) TYPE: amino acid				
	(C) STRANDEDNESS: not relevant				
	(D) TOPOLOGY: not relevant				
	(ii) MOLECULE TYPE: peptide				
45	(iii) HYPOTHETICAL: NO				
	(iv) ANTI-SENSE: NO				
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:				
	Met Val Gln Gln Ala Ala Thr Val Ser Leu Met Val Leu Phe Leu Val				
		5		10	15
55	Pro Gln Leu Arg Asn Ala Tyr Gly Thr Ala Ala Ile Gly Ile Ile Cys				
		20		25	30
60	Gly Leu Tyr Trp Ala Val Ser Ser Asn Met Thr Val Glu Ala Thr Gln				
		35		40	45

WO 98/26072

PCT/US97/22578

-230-

Arg Leu Thr Gly Gly Gly Phe Ala Ile Gly His Gln Gln Gln Phe  
 50 55 60  
 5 Ala Ile Trp Phe Val Asp Lys Val Ala Gly Arg Phe Gly Lys Lys Glu  
 65 70 75  
 Glu Ser Leu Asp Asn Leu Lys Leu Pro Lys Phe Leu Ser Ile Phe His  
 85 90 95  
 10 Asp Thr Val Val Ala Ser Ala Thr Leu Met Leu Val Phe Phe Gly Ala  
 100 105 110  
 Ile Leu Leu Ile Leu Gly Pro Asp Ile Met Ser Asn Lys Glu Val Ile  
 115 120 125  
 15 Thr Ser Gly Thr Leu Phe Asn Pro Ala Lys Gln Asp Phe Phe Met Tyr  
 130 135 140  
 Ile Ile Gln Thr Ala Phe Thr Phe Ser Val Tyr Leu Phe Val Leu Met  
 145 150 155 160  
 Gln Gly Val Arg Met Phe Val Ser Glu Leu Thr Asn Ala Phe Gln Gly  
 165 170 175  
 25 Ile Ser Asn Lys Leu Leu Pro Gly Ser Phe Pro Ala Val Asp Val Ala  
 180 185 190  
 Ala Ser Tyr Gly Phe Gly Ser Pro Asn Ala Val Leu Ser Gly Phe Thr  
 195 200 205  
 30 Phe Gly Leu Ile Gly Gln Leu Ile Thr Ile Val Leu Leu Ile Val Phe  
 210 215 220  
 Lys Asn Pro Ile Leu Ile Ile Thr Gly Phe Val Pro Val Phe Phe Asp  
 225 230 235 240  
 35 Asn Ala Ala Ile Ala Val Tyr Ala Asp Lys Arg Gly Gly Trp Lys Ala  
 245 250 255  
 40 Ala Val Ile Leu Ser Phe Ile Ser Gly Val Leu Gln Val Ala Leu Gly  
 260 265 270  
 Ala Leu Cys Val Ala Leu Leu Asp Leu Ala Ser Tyr Gly Gly Tyr His  
 275 280 285  
 45 Gly Asn Ile Asp Phe Glu Phe Pro Trp Leu Gly Phe Gly Tyr Ile Phe  
 290 295 300  
 Lys Tyr Leu Gly Ile Val Gly Tyr Val Leu Val Cys Leu Phe Leu Leu  
 305 310 315 320  
 Val Ile Pro Gln Leu Gln Phe Ala Lys Ala Lys Asp Lys Glu Lys Tyr  
 325 330 335  
 55 Tyr Asn Gly Glu Val Gln Glu Glu Ala  
 340 345

(2) INFORMATION FOR SEQ ID NO:144:

60 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 287 amino acids

WO 98/26072

PCT/US97/22578

-231-

(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

5 (ii) MOLECULE TYPE: peptide  
(iii) HYPOTHETICAL: NO  
10 (iv) ANTI-SENSE: NO  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:  
15 Met Val Arg Pro Ile Gly Ile Tyr Glu Lys Ala Thr Pro Thr His Phe  
1 5 10 15  
20 Thr Trp Leu Glu Arg Leu Asn Phe Ala Lys Glu Leu Gly Phe Asp Phe  
20 25 30  
Val Glu Met Ser Ile Asp Glu Arg Asp Glu Arg Leu Ala Arg Leu Asp  
35 40 45  
25 Trp Ser Lys Glu Glu Arg Leu Glu Val Val Lys Ala Ile Tyr Glu Thr  
50 55 60  
Gly Val Arg Ile Pro Ser Ile Cys Phe Ser Gly His Arg Arg Tyr Pro  
65 70 75 80  
30 Leu Gly Ser Lys Asp Pro Val Leu Glu Glu Lys Ser Leu Glu Leu Met  
85 90 95  
35 Lys Lys Cys Ile Glu Leu Ala Gln Asp Leu Gly Val Arg Thr Ile Gln  
100 105 110  
Leu Ala Gly Tyr Asp Val Tyr Tyr Glu Glu Lys Ser Pro Gln Thr Arg  
115 120 125  
40 Gln Arg Phe Ile Lys Asn Leu Arg Lys Ala Cys Asp Trp Ala Glu Glu  
130 135 140  
Ala Gln Val Val Leu Ala Ile Glu Ile Met Asp Asp Pro Phe Ile Asn  
145 150 155 160  
45 Ser Ile Glu Lys Tyr Leu Ala Ile Glu Lys Glu Ile Asp Ser Pro Phe  
165 170 175  
50 Leu Phe Val Tyr Pro Asp Ile Gly Asn Val Ser Ala Trp His Asn Asp  
180 185 190  
Ile Tyr Ser Glu Phe Tyr Leu Gly His His Ala Ile Ala Ala Leu His  
195 200 205  
55 Leu Lys Asp Thr Tyr Ala Val Thr Glu Ser Ser Lys Gly Gln Phe Arg  
210 215 220  
Asp Val Pro Phe Gly Gln Gly Cys Val Lys Trp Glu Glu Ala Phe Asp  
225 230 235 240  
60 Ile Leu Lys Glu Thr Asn Tyr Asn Gly Pro Phe Leu Ile Glu Met Trp

PCT/US97/22578

```

      245                      250                      255
Ser Glu Asn Cys Glu Thr Val Glu Glu Thr Arg Ala Ala Val Gln Glu
      260                      265                      270
Ala Gln Ala Phe Leu Tyr Pro Leu Ile Lys Lys Ala Gly Leu Met
      275                      280                      285

(2) INFORMATION FOR SEQ ID NO:145:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 221 amino acids
    (B) TYPE: amino acid
    (C) STRANDEDNESS: not relevant
    (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:
Met Thr Lys Arg Ile Pro Asn Leu Gln Val Ala Leu Asp His Ser Asp
 1                      5                      10                      15
Leu Gln Gly Ala Ile Lys Ala Ala Val Ser Val Gly Gln Glu Val Asp
      20                      25
Ile Ile Glu Ala Gly Thr Val Cys Leu Leu Gln Val Gly Ser Glu Leu
      35                      40                      45
Ala Glu Val Leu Arg Ser Leu Phe Pro Asp Lys Ile Ile Val Ala Asp
      50                      55                      60
Thr Lys Cys Ala Asp Ala Gly Gly Thr Val Ala Lys Asn Asn Ala Val
      65                      70                      75                      80
Arg Gly Ala Asp Trp Met Thr Cys Ile Cys Cys Ala Thr Ile Pro Thr
      85                      90                      95
Met Glu Ala Ala Leu Lys Ala Ile Lys Thr Glu Arg Gly Glu Arg Gly
      100                      105                      110
Glu Ile Gln Ile Glu Leu Tyr Gly Asp Trp Thr Phe Glu Gln Ala Gln
      115                      120                      125
Leu Trp Leu Asp Ala Gly Ile Ser Gln Ala Ile Tyr His Gln Ser Arg
      130                      135                      140
Asp Ala Leu Leu Ala Gly Glu Thr Trp Gly Glu Lys Asp Leu Asn Lys
      145                      150                      155                      160
Val Lys Lys Leu Ile Asp Met Gly Phe Arg Val Ser Val Thr Gly Gly
      165                      170                      175
Leu Asp Val Asp Thr Leu Lys Leu Phe Glu Gly Val Asp Val Phe Thr
      180                      185                      190

```

WO 98/26072

PCT/US97/22578

-233-

Phe Ile Ala Gly Arg Gly Ile Thr Glu Ala Ala Asp Pro Ala Gly Ala  
195 200 205

5 Ala Arg Ala Phe Lys Asp Glu Ile Lys Arg Ile Trp Gly  
210 215 220

(2) INFORMATION FOR SEQ ID NO:146:

(i) SEQUENCE CHARACTERISTICS:  
10 (A) LENGTH: 161 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

15 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

20 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

25 Met Asn Leu Lys Gln Ala Leu Ile Asp Asn Asp Ser Ile Arg Leu Gly  
1 5 10 15

Leu Glu Ala Asn Asn Trp Lys Glu Ala Val Lys Val Ala Val Asp Pro  
20 25 30

30 Leu Ile Glu Ser Gly Ala Ile Leu Pro Glu Tyr Tyr Asp Ala Ile Ile  
35 40 45

Glu Ser Thr Glu Glu Tyr Gly Pro Tyr Tyr Ile Leu Met Pro Gly Met  
50 55 60

35 Ala Met Pro His Ala Arg Pro Glu Ala Gly Val Gln Ser Asp Ala Phe  
65 70 75 80

40 Ser Leu Ile Thr Leu Gln Asn Pro Val Val Phe Ser Asp Gly Lys Glu  
85 90 95

Val Ser Val Leu Leu Ala Leu Ala Ala Thr Ser Ser Lys Ile His Thr  
100 105 110

45 Ser Val Ala Ile Pro Gln Ile Ile Ala Leu Phe Glu Leu Glu Asp Ser  
115 120 125

Ile Ala Arg Leu Gln Ala Cys Gln Thr Lys Glu Asp Val Leu Ala Met  
130 135 140

50 Ile Glu Glu Ser Lys Asp Ser Pro Tyr Leu Glu Gly Leu Asp Leu Glu  
145 150 155 160

55 Ser

(2) INFORMATION FOR SEQ ID NO:147:

(i) SEQUENCE CHARACTERISTICS:  
60 (A) LENGTH: 293 amino acids  
(B) TYPE: amino acid

-234-

(C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

Met	Ser	Arg	Asp	Ile	Ile	Lys	Leu	Asp	Gln	Ile	Asp	Val	Thr	Phe	His
1				5					10					15	
Gln	Lys	Lys	Arg	Thr	Ile	Thr	Ala	Val	Lys	Asp	Val	Thr	Ile	His	Ile
			20				25					30			
Gln	Glu	Gly	Asp	Ile	Tyr	Gly	Ile	Val	Gly	Tyr	Ser	Gly	Ala	Gly	Lys
	35					40					45				
Ser	Thr	Leu	Val	Arg	Val	Ile	Asn	Leu	Leu	Gln	Lys	Pro	Ser	Ala	Gly
	50				55					60					
Lys	Ile	Thr	Ile	Asp	Asp	Val	Ile	Phe	Asp	Gly	Lys	Val	Thr	Leu	
	65			70				75						80	
Thr	Ala	Glu	Gln	Leu	Arg	Arg	Lys	Arg	Gln	Asp	Ile	Gly	Met	Ile	Phe
				85				90					95		
Gln	His	Phe	Asn	Leu	Met	Ser	Gln	Lys	Thr	Ala	Glu	Glu	Asn	Val	Ala
			100				105						110		
Phe	Ala	Leu	Lys	His	Ser	Gly	Leu	Ser	Lys	Glu	Glu	Lys	Lys	Ala	Lys
	115						120					125			
Val	Ala	Lys	Leu	Leu	Asp	Leu	Val	Gly	Leu	Ala	Asp	Arg	Ala	Glu	Asn
	130					135					140				
Tyr	Pro	Ser	Gln	Leu	Ser	Gly	Gly	Gln	Lys	Gln	Arg	Val	Ala	Ile	Ala
	145			150					155					160	
Arg	Ala	Leu	Ala	Asn	Asp	Pro	Lys	Ile	Leu	Ile	Ser	Asp	Glu	Ser	Thr
		165					170					175			
Ser	Ala	Leu	Asp	Pro	Lys	Thr	Thr	Lys	Gln	Ile	Leu	Ala	Leu	Leu	Gln
	180						185					190			
Asp	Leu	Asn	Gln	Lys	Leu	Gly	Leu	Thr	Val	Val	Leu	Ile	Thr	His	Glu
	195					200					205				
Met	Gln	Ile	Val	Lys	Asp	Ile	Ala	Asn	Arg	Val	Ala	Val	Met	Gln	Asp
	210					215					220				
Gly	His	Leu	Ile	Glu	Glu	Gly	Ser	Val	Leu	Glu	Ile	Phe	Ser	Asn	Pro
	225				230				235					240	
Lys	Gln	Pro	Leu	Thr	Gln	Asp	Phe	Ile	Ser	Thr	Ala	Thr	Gly	Ile	Asp
		245					250						255		
Glu	Ala	Met	Val	Lys	Ile	Glu	Lys	Gln	Glu	Ile	Val	Glu	His	Leu	Ser





WO 98/26072

PCT/US97/22578

-236-

Lys Asp Val Val Leu Asn Gln His Pro Gln Asp Ile Asp Ser Leu Glu  
195 200 205

Tyr

5

(2) INFORMATION FOR SEQ ID NO:149:

(i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 213 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

15

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

25

Met Ile Glu Leu Lys Asn Ile Thr Lys Thr Ile Gly Gly Lys Val Ile  
1 5 10 15

30

Leu Asp Asn Leu Ser Leu Arg Ile Asp Gln Gly Asp Leu Val Ala Ile  
20 25 30

Val Gly Lys Ser Gly Ser Gly Lys Ser Thr Leu Leu Asn Leu Leu Gly  
35 40 45

35

Leu Ile Asp Gly Asp Tyr Ser Gly Arg Tyr Glu Ile Phe Gly Gln Thr  
50 55 60

40

Asn Leu Ala Val Asn Ser Ala Lys Ser Gln Thr Ile Ile Arg Glu His  
65 70 75 80

Ile Ser Tyr Leu Phe Gln Asn Phe Ala Leu Ile Asp Asp Glu Thr Val  
85 90 95

45

Glu Tyr Asn Leu Met Leu Ala Leu Lys Tyr Val Lys Leu Pro Lys Lys  
100 105 110

Asp Lys Leu Lys Lys Val Glu Glu Ile Leu Glu Arg Val Gly Leu Ser  
115 120 125

50

Ala Thr Leu His Gln Arg Val Ser Glu Leu Ser Gly Gly Glu Gln Gln  
130 135 140

Arg Ile Ala Val Ala Arg Ala Ile Leu Lys Pro Ser Gln Leu Ile Leu  
145 150 155 160

55

Ala Asp Glu Pro Thr Gly Ser Leu Asp Pro Glu Asn Arg Asp Leu Val  
165 170 175

60

Leu Lys Phe Leu Leu Glu Met Asn Arg Glu Gly Lys Thr Val Ile Ile  
180 185 190

-237-

Val Thr His Asp Ala Tyr Val Ala Gln Gln Cys His Arg Val Ile Glu  
 195 200 205

Leu Gly Glu Gly Lys  
 210

## (2) INFORMATION FOR SEQ ID NO:150:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 84 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

Ala Lys Pro Lys Gln Leu Ser Gly Gly Gln Lys Gln Arg Val Ala Ile  
 1 5 10 15

Ala Arg Ala Leu Ser Met Asn Pro Asp Ala Ile Leu Phe Asp Glu Pro  
 20 25 30

Thr Ser Ala Leu Asp Pro Glu Met Val Gly Glu Val Leu Lys Ile Met  
 35 40 45

Gln Asp Leu Ala Gln Glu Gly Leu Thr Met Ile Val Val Thr His Glu  
 50 55 60

Met Glu Phe Ala Arg Asp Val Ser His Arg Val Ile Phe Met Asp Lys  
 65 70 75 80

Gly Val Ile Pro

## (2) INFORMATION FOR SEQ ID NO:151:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 118 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

Tyr Tyr Gly Asp Tyr His Ala Leu Arg Asn Ile Asn Leu Arg Phe Glu  
 1 5 10 15

-238-

Lys Gly Gln Val Val Val Leu Leu Gly Pro Ser Gly Ser Gly Lys Ser  
 20 25 30  
 5 Thr Leu Ile Arg Thr Ile Asn Gly Leu Glu Ala Val Asp Lys Gly Ser  
 35 40 45  
 Leu Leu Val Asn Gly His Gln Val Ala Gly Ala Ser Gln Lys Asp Leu  
 50 55 60  
 10 Val Pro Leu Arg Lys Glu Val Gly Met Val Phe Gln His Phe Asn Leu  
 65 70 75 80  
 Tyr Pro His Lys Thr Val Leu Glu Asn Val Thr Leu Ala Pro Ile Lys  
 85 90 95  
 15 Val Leu Gly Ile Asp Lys Lys Glu Ala Glu Lys Thr Ala Gln Lys Tyr  
 100 105 110  
 20 Leu Glu Phe Val Asn Met  
 115

## (2) INFORMATION FOR SEQ ID NO:152:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 237 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant  
 30 (ii) MOLECULE TYPE: peptide  
 (iii) HYPOTHETICAL: NO  
 35 (iv) ANTI-SENSE: NO  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:  
 40 Met Thr Lys Lys Gln Leu His Leu Val Ile Val Thr Gly Met Gly Gly  
 1 5 10 15  
 Ala Gly Lys Thr Val Ala Ile Gln Ser Phe Glu Asp Leu Gly Tyr Phe  
 20 25 30  
 45 Thr Ile Asp Asn Met Pro Pro Ala Leu Leu Pro Lys Phe Leu Gln Leu  
 35 40 45  
 Val Glu Ile Lys Glu Asp Asn Pro Lys Leu Ala Leu Val Val Asp Met  
 50 55 60  
 Arg Ser Arg Ser Phe Phe Ser Glu Ile Gln Ala Val Leu Asp Glu Leu  
 65 70 75 80  
 55 Glu Asn Gln Asp Gly Leu Asp Phe Lys Ile Leu Phe Leu Asp Ala Ala  
 85 90 95  
 Asp Lys Glu Leu Val Ala Arg Tyr Lys Glu Thr Arg Arg Ser His Pro  
 100 105 110  
 60 Leu Ala Ala Asp Gly Arg Ile Leu Asp Gly Ile Lys Leu Glu Arg Glu  
 115 120 125

-239-

Leu Leu Ala Pro Leu Lys Asn Met Ser Gln Asn Val Val Asp Thr Thr  
 130 135 140  
 5 Glu Leu Thr Pro Arg Glu Leu Arg Lys Thr Leu Ala Glu Gln Phe Ser  
 145 150 155 160  
 Asp Gln Glu Gln Ala Gln Ser Phe Arg Ile Glu Val Met Ser Phe Gly  
 165 170 175  
 10 Phe Lys Tyr Gly Ile Pro Ile Asp Ala Asp Leu Val Phe Asp Val Arg  
 180 185 190  
 Phe Leu Pro Asn Pro Tyr Tyr Leu Pro Glu Leu Arg Asn Gln Thr Gly  
 195 200 205  
 15 Val Asp Glu Pro Val Tyr Asp Tyr Val Met Asn His Pro Glu Ser Glu  
 210 215 220  
 20 Asp Phe Tyr Gln His Leu Leu Ala Leu Ile Glu Pro Ile  
 225 230 235

## (2) INFORMATION FOR SEQ ID NO:153:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 180 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant  
 30 (ii) MOLECULE TYPE: peptide  
 (iii) HYPOTHETICAL: NO  
 35 (iv) ANTI-SENSE: NO  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:  
 40 Met Leu Glu Asn Asp Ile Lys Lys Val Leu Val Ser His Asp Glu Ile  
 1 5 10 15  
 Thr Glu Ala Ala Lys Lys Leu Gly Ala Gln Leu Thr Lys Asp Tyr Ala  
 20 25 30  
 45 Gly Lys Asn Pro Ile Leu Val Gly Ile Leu Lys Gly Ser Ile Pro Phe  
 35 40 45  
 50 Met Ala Glu Leu Val Lys His Ile Asp Thr His Ile Glu Met Asp Phe  
 50 55 60  
 Met Met Val Ser Ser Tyr His Gly Gly Thr Ala Ser Ser Gly Val Ile  
 65 70 75 80  
 55 Asn Ile Lys Gln Asp Val Thr Gln Asp Ile Lys Gly Arg His Val Leu  
 85 90 95  
 Phe Val Glu Asp Ile Ile Asp Thr Gly Gln Thr Leu Lys Asn Leu Arg  
 100 105 110  
 60 Asp Met Phe Lys Glu Arg Glu Ala Ala Ser Val Lys Ile Ala Thr Leu

-240-

115                                      120                                      125  
 Leu Asp Lys Pro Glu Gly Arg Val Val Glu Ile Glu Ala Asp Tyr Thr  
 130                                      135                                      140  
 5                                      Cys Phe Thr Ile Pro Asn Glu Phe Val Val Gly Tyr Gly Leu Asp Tyr  
 145                                      150                                      155                                      160  
 10                                      Lys Glu Asn Tyr Arg Asn Leu Pro Tyr Ile Gly Val Leu Lys Glu Glu  
 165                                      170                                      175  
 Val Tyr Ser Asn  
 180  
 15 (2) INFORMATION FOR SEQ ID NO:154:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 193 amino acids  
 (B) TYPE: amino acid  
 20 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant  
 (ii) MOLECULE TYPE: peptide  
 25 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:  
 Met Lys Ile Gly Ile Leu Ala Leu Gln Gly Ala Phe Ala Glu His Ala  
 1                                      5                                      10                                      15  
 35 Lys Val Leu Asp Gln Leu Gly Val Glu Ser Val Glu Leu Arg Asn Leu  
 20                                      25                                      30  
 Asp Asp Phe Gln Gln Asp Gln Ser Asp Leu Ser Gly Leu Ile Leu Pro  
 35                                      40                                      45  
 40 Gly Gly Glu Ser Thr Thr Met Gly Lys Leu Leu Arg Asp Gln Asn Met  
 50                                      55                                      60  
 45 Leu Leu Pro Ile Arg Glu Ala Ile Leu Ser Gly Leu Pro Val Phe Gly  
 65                                      70                                      75                                      80  
 Thr Cys Ala Gly Leu Ile Leu Leu Ala Lys Glu Ile Thr Ser Gln Lys  
 85                                      90                                      95  
 50 Glu Ser His Leu Gly Thr Met Asp Met Val Val Glu Arg Asn Ala Tyr  
 100                                      105                                      110  
 55 Gly Arg Gln Leu Gly Ser Phe Tyr Thr Glu Ala Glu Cys Lys Gly Val  
 115                                      120                                      125  
 Gly Lys Ile Pro Met Thr Phe Ile Arg Gly Pro Ile Ile Ser Ser Val  
 130                                      135                                      140  
 60 Gly Glu Gly Val Glu Ile Leu Ala Ile Val Asn Asn Gln Ile Val Ala  
 145                                      150                                      155                                      160

-241-

Ala Gln Glu Lys Asn Met Leu Val Ser Ser Phe His Pro Glu Leu Thr  
165 170 175

5 Asp Asp Val Arg Leu His Gln Tyr Phe Ile Asn Met Cys Lys Glu Lys  
180 185 190

Ser

10 (2) INFORMATION FOR SEQ ID NO:155:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 189 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

15 (ii) MOLECULE TYPE: peptide

20 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

Glu Ser Glu Val Leu Ser Pro Ala Asp Asp Arg Phe His Val Asp Lys  
1 5 10 15

Lys Glu Phe Gln Val Pro Phe Val Cys Gly Ala Lys Asp Leu Gly Glu  
20 25 30

Ala Leu Arg Arg Ile Ala Glu Gly Ala Ser Met Ile Arg Thr Lys Gly  
35 35 40 45

Glu Pro Gly Thr Gly Asp Ile Val Gln Ala Val Arg His Met Arg Met  
50 55 60

40 Met Asn Gln Glu Ile Arg Arg Ile Gln Asn Leu Arg Glu Asp Glu Leu  
65 70 75 80

Tyr Val Ala Ala Lys Asp Leu Gln Val Pro Val Glu Leu Val Gln Tyr  
85 90 95

45 Val His Glu His Gly Lys Leu Pro Val Val Asn Phe Ala Ala Gly Gly  
100 105 110

Val Ala Thr Pro Ala Asp Ala Ala Leu Met Met Gln Leu Gly Ala Glu  
115 120 125

Gly Val Phe Val Gly Ser Gly Ile Phe Lys Ser Gly Asp Pro Val Lys  
130 135 140

55 Arg Ala Ser Ala Ile Val Lys Ala Val Thr Asn Phe Arg Asn Pro Gln  
145 150 155 160

Ile Leu Ala Gln Ile Ser Glu Asp Leu Gly Glu Ala Met Val Gly Ile  
165 170 175

60 Asn Glu Asn Ile Gln Ile Leu Met Ala Glu Arg Gly Lys

-242-

180

185

## (2) INFORMATION FOR SEQ ID NO:156:

5

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 162 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

10

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

15

(iv) ANTI-SENSE: NO

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

Asp Lys Gly Trp Phe Val Leu Gln Thr Tyr Ser Gly Tyr Glu Asn Lys  
 1 5 10 15

25

Val Lys Glu Asn Leu Leu Gln Arg Ala Gln Thr Tyr Asn Met Leu Asp  
 20 25 30

30

Asn Ile Leu Arg Val Glu Ile Pro Thr Gln Thr Val Gln Val Glu Lys  
 35 40 45

35

Asn Gly Lys Arg Lys Glu Val Glu Glu Asn Arg Phe Pro Gly Tyr Val  
 50 55 60

40

Leu Val Glu Met Val Met Thr Asp Glu Ala Trp Phe Val Val Arg Asn  
 65 70 75 80

45

Ala Gln Ser Pro Thr Lys Phe Ile Ser Glu Gln Thr Ala Tyr Glu Ile  
 85 90 95

50

Asp Glu Glu Val Arg Ser Leu Leu Asn Glu Ala Arg Asn Lys Ala Ala  
 100 105 110

55

Glu Ile Ile Gln Ser Asn Arg Glu Thr His Lys Leu Ile Ala Glu Ala  
 115 120 125

60

Leu Leu Lys Tyr Glu Thr Leu Asp Ser Thr Gln Ile Lys Ala Leu Tyr  
 130 135 140

65

Glu Thr Gly Lys Met Pro Glu Ser Ser Arg Arg Gly Ile Ser Cys Thr  
 145 150 155 160

Ile Leu

70

## (2) INFORMATION FOR SEQ ID NO:157:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 208 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

75





WO 98/26072

PCT/US97/22578

-244-

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:

10 Gln Gln Ser Val Lys Lys Lys Val Leu Pro Ala Ile Glu Arg Arg Ile  
 1 5 10 15  
 Arg Thr Glu Leu Thr Glu Lys Ala Glu Gly Ala Ile Gln Leu Phe  
 20 25 30  
 15 Ser Asp Asn Leu Arg Asn Leu Leu Leu Val Ala Pro Leu Lys Gly Arg  
 35 40  
 Val Val Leu Glu Phe Asp Pro Ala Phe Arg Thr Gly Ala Lys Leu Ala  
 50 55 60  
 20 Val Val Asp Ala Thr Gly Lys Met Leu Thr Gln Val Ile Tyr Pro  
 65 70 75 80  
 Val Lys Pro Ala Ser Ala Arg Gln Ile Glu Glu Ala Lys Lys Asp Leu  
 85 90 95  
 25 Ala Asp Leu Ile Gly Gln Tyr Gly Val Glu Ile Ile Ala Ile Gly Asn  
 100 105 110  
 30 Gly Thr Ala Ser Arg Glu Ser Glu Ala Phe Val Ala Glu Val Leu Lys  
 115 120 125  
 Asp Phe Pro Glu Val Ser Tyr Val Ile Val Asn Glu Ser Gly Ala Ser  
 130 135 140  
 35 Val Tyr Ser Ala Ser Glu Leu Ala Arg Gln Glu Phe Pro Asp Leu Thr  
 145 150 155 160  
 Val Glu Lys Arg Ser Ala Ile Ser Ile Ala Arg Arg Leu Gln Asp Pro  
 165 170 175  
 40 Leu Ala Glu Leu Val Lys Ile Asp Pro Lys Ser Ile Gly Val Gly Gln  
 180 185 190  
 45 Tyr Gln His Asp Val Ser Gln Lys Lys Leu Ser Glu Ser Leu Asp Phe  
 195 200 205  
 Val Val Asp Thr Val Val Asn Gln Val Gly Val Asn Val Asn Thr Ala  
 210 215 220  
 50 Ser Pro Ala Leu Leu Ser His Val Ala Gly Leu Asn Lys Thr Ile Ser  
 225 230 235 240  
 Glu Asn Ile Val Lys Tyr Arg Glu Glu Glu Gly Lys Ile Thr Ser Arg  
 245 250 255  
 55 Ala Gln Ile Lys Lys Val Pro Arg Leu Gly Ala Lys Ala Phe Glu Gln  
 260 265 270  
 60 Ala Ala Gly Phe Leu Arg Ile Pro Glu Ser Ser Asn Ile Leu Asp Asn  
 275 280 285

WO 98/26072

PCT/US97/22578

-245-

Thr Gly Val His Pro Glu Asn Tyr Thr Ala Val Lys Leu Phe Lys Arg  
290 295 300

5 Leu Asp Ile Lys Asp Leu Asn Glu Glu Ala Ser Lys Leu Lys Ser Leu  
305 310 315 320

Ser Val Lys Glu Met Ala Gln Glu Leu Asp Leu Gly Pro Glu Thr Leu  
325 330 335

10 Lys Asp Ile Ile Ala Asp Leu Leu Lys Pro Gly Arg Asp Phe Arg Asp  
340 345 350

15 Ser Phe Asp Ala Pro Val Leu Arg Gln Asp Val Leu Asp Ile Lys Asp  
355 360 365

Leu Val Val Gly Gln Lys Leu Glu Gly Val Val Arg Asn Val Val Asp  
370 375 380

20 Phe Gly Ala Phe Val Asp Ile Gly Val His Glu Asp Gly Leu Ile His  
385 390 395 400

Ile Leu Ile

25 (2) INFORMATION FOR SEQ ID NO:159:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 179 amino acids

30 (B) TYPE: amino acid

(C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

35 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:

45 Met Phe Arg Ala Ala Met Ala Asn Gln Thr Glu Met Gly Val Leu Ala  
1 5 10 15

Lys Ser Tyr Ile Asp Lys Gly Glu Leu Val Pro Asp Glu Val Thr Asn  
20 25 30

50 Gly Ile Val Lys Glu Arg Leu Ser Gln Asp Asp Ile Lys Glu Thr Gly  
35 40 45

55 Phe Leu Leu Asp Gly Tyr Pro Arg Thr Ile Glu Gln Ala His Ala Leu  
50 55 60

Asp Lys Thr Leu Ala Glu Leu Gly Ile Glu Leu Glu Gly Ile Ile Asn  
65 70 75 80

60 Ile Glu Val Asn Pro Asp Ser Leu Leu Glu Arg Leu Ser Gly Arg Ile  
85 90 95

WO 98/26072

PCT/US97/22578

-246-

Ile His Arg Val Thr Gly Glu Thr Phe His Lys Val Phe Asn Pro Pro  
 100 105 110

5 Val Asp Tyr Lys Glu Glu Asp Tyr Tyr Gln Arg Glu Asp Asp Lys Pro  
 115 120 125

Glu Thr Val Lys Arg Arg Leu Asp Val Asn Ile Ala Gln Gly Glu Pro  
 130 135 140

10 Ile Ile Ala His Tyr Arg Ala Lys Gly Leu Val His Asp Ile Glu Gly  
 145 150 155 160

Asn Gln Asp Ile Asn Asp Val Phe Ser Asp Ile Glu Lys Val Leu Thr  
 165 170 175

15 Asn Leu Lys

20 (2) INFORMATION FOR SEQ ID NO:160:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:

40 Met Ile Glu Phe Glu Lys Pro Asn Ile Thr Lys Ile Asp Glu Asn Lys  
 1 5 10 15

Asp Tyr Gly Lys Phe Val Ile Glu Pro Leu Glu Arg Gly Tyr Gly Thr  
 20 25 30

45 Thr Leu Gly Asn Ser Leu Arg Arg Val Leu Leu Ala Ser Leu Pro Gly  
 35 40 45

Ala Ala Val Thr Ser Ile Asn Ile Asp Gly Val Leu His Glu Phe Asp  
 50 55 60

50 Thr Val Pro Gly Val Arg Glu Asp Val Met Gln Ile Ile Leu Asn Ile  
 65 70 75 80

55 Lys Gly Ile Ala Val Lys Ser Tyr Val Glu Asp Glu Lys Ile Ile Glu  
 85 90 95

Leu Asp Val Glu Gly Pro Ala Glu Val Thr Ala Gly Asp Ile Leu Thr  
 100 105 110

60 Asp Ser Asp Ile Glu Ile Val Asn Pro Asp His Tyr Leu Phe Thr Ile  
 115 120 125

WO 98/26072

PCT/US97/22578

-247-

Gly Glu Gly Ser Ser Leu Lys Ala Thr Met Thr Val Asn Ser Gly Arg  
130 135 140

5 Gly Tyr Val Pro Ala Asp Glu Asn Lys Lys Asp Asn Ala Pro Val Gly  
145 150 155 160

Thr Leu Ala Val Asp Ser Ile Tyr Thr Pro Val Thr Lys Val Asn Tyr  
165 170 175

10 Gln Val Glu Pro Ala Arg Val Gly Ser Asn Asp Gly Phe Asp Ser  
180 185 190

(2) INFORMATION FOR SEQ ID NO:161:

15 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 335 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
20 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

25 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:161:

30 Glu Tyr Leu Gly Ala Thr Val Gln Val Ile Pro His Ile Thr Asp Ala  
1 5 10 15

35 Leu Lys Glu Lys Ile Lys Ser Ala Ala Leu Thr Thr Asp Ser Asp Val  
20 25 30

Ile Ile Thr Glu Val Gly Gly Thr Val Gly Asp Ile Glu Ser Leu Pro  
35 40 45

40 Phe Leu Glu Ala Leu Arg Gln Met Lys Ala Asp Val Gly Ala Asp Asn  
50 55 60

Val Met Tyr Ile His Thr Thr Leu Pro Tyr Leu Lys Ala Ala Gly Glu  
65 70 75 80

45 Met Lys Lys Pro Thr Gln His Ser Val Lys Leu Arg Gly Leu Gly Ile  
85 90 95

Gln Pro Asn Met Leu Val Ile Arg Thr Glu Glu Pro Ala Gly Gln Gly  
100 105 110

Ile Lys Asn Lys Leu Ala Gln Phe Cys Asp Val Ala Pro Glu Ser Leu  
115 120 125

55 Ile Glu Ser Leu Asp Val Glu His Leu Tyr Gln Ile Pro Leu Asn Leu  
130 135 140

Gln Ala Gln Gly Met Asp Gln Ile Val Cys Asp His Leu Lys Leu Asp  
145 150 155 160

60 Ala Pro Ala Ala Asp Met Thr Glu Trp Ser Ala Met Val Asp Lys Val

WO 98/26072

PCT/US97/22578

-248-

		165		170		175	
		Met Asn Leu Lys Lys Gln Val Lys Ile Ser Leu Val Gly Lys Tyr Val					
		180		185		190	
5		Glu Leu Gln Asp Ala Tyr Ile Ser Val Val Glu Ala Leu Lys His Ser					
		195		200		205	
10		Gly Tyr Val Asn Asp Val Glu Val Lys Ile Asn Trp Val Asn Ala Asn					
		210		215		220	
		Asp Val Thr Ala Glu Asn Val Ala Glu Leu Leu Ser Asp Ala Asp Gly					
		225		230		235	
15		Ile Ile Val Pro Gly Gly Phe Gly Gln Arg Gly Thr Glu Gly Lys Ile					
		245		250		255	
		Gln Ala Ile Arg Tyr Ala Arg Glu Asn Asp Val Pro Met Leu Gly Val					
		260		265		270	
20		Cys Leu Gly Met Gln Leu Thr Cys Ile Glu Phe Ala Arg His Val Leu					
		275		280		285	
25		Gly Leu Glu Gly Ala Asn Ser Ala Glu Leu Ala Pro Glu Thr Lys Tyr					
		290		295		300	
		Pro Ile Ile Asp Ile Met Arg Asp Gln Ile Asp Ile Glu Asp Met Gly					
		305		310		315	
30		Gly Thr Leu Arg Leu Gly Leu Tyr Pro Ser Lys Leu Lys Arg Leu					
		325		330		335	

(2) INFORMATION FOR SEQ ID NO:162:

35	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 301 amino acids
	(B) TYPE: amino acid
	(C) STRANDEDNESS: not relevant
	(D) TOPOLOGY: not relevant
40	(ii) MOLECULE TYPE: peptide
	(iii) HYPOTHETICAL: NO
45	(iv) ANTI-SENSE: NO
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:
50	Met Ser Glu Lys Leu Val Glu Ile Lys Asp Leu Glu Ile Ser Phe Gly
	1 5 10 15
	Glu Gly Ser Lys Lys Phe Val Ala Val Lys Asn Ala Asn Phe Phe Ile
	20 25 30
55	Asn Lys Gly Glu Thr Phe Ser Leu Val Gly Glu Ser Gly Ser Gly Lys
	35 40 45
60	Thr Thr Ile Gly Arg Ala Ile Ile Gly Leu Asn Asp Thr Ser Asn Gly
	50 55 60

WO 98/26072

PCT/US97/22578

-249-

	Asp	Ile	Ile	Phe	Asp	Gly	Gln	Lys	Ile	Asn	Gly	Lys	Lys	Ser	Arg	Glu
	65					70					75					80
5	Gln	Ala	Ala	Glu	Leu	Ile	Arg	Arg	Ile	Gln	Met	Ile	Phe	Gln	Asp	Pro
					85					90					95	
	Ala	Ala	Ser	Leu	Asn	Glu	Arg	Ala	Thr	Val	Asp	Tyr	Ile	Ile	Ser	Glu
				100					105						110	
10	Gly	Leu	Tyr	Asn	His	Arg	Leu	Phe	Lys	Asp	Glu	Glu	Glu	Arg	Lys	Glu
			115				120						125			
	Lys	Val	Gln	Ser	Ile	Ile	Arg	Glu	Val	Gly	Leu	Leu	Ala	Glu	His	Leu
			130				135						140			
15	Thr	Arg	Tyr	Pro	His	Glu	Phe	Ser	Gly	Gly	Gln	Arg	Gln	Arg	Ile	Gly
						150					155					160
	Ile	Ala	Arg	Ala	Leu	Val	Met	Gln	Pro	Asp	Phe	Val	Ile	Ala	Asp	Glu
20					165					170						175
	Pro	Ile	Ser	Ala	Leu	Asp	Val	Ser	Val	Arg	Ala	Gln	Val	Leu	Asn	Leu
					180					185					190	
25	Leu	Lys	Lys	Phe	Gln	Lys	Glu	Leu	Gly	Leu	Thr	Tyr	Leu	Phe	Ile	Ala
				195					200					205		
	His	Asp	Leu	Ser	Val	Val	Arg	Phe	Ile	Ser	Asp	Arg	Ile	Ala	Val	Ile
		210					215					220				
30	Tyr	Lys	Gly	Val	Ile	Val	Glu	Val	Ala	Glu	Thr	Glu	Glu	Leu	Phe	Asn
						230					235					240
	Asn	Pro	Ile	His	Pro	Tyr	Thr	Gln	Ala	Leu	Leu	Ser	Ala	Val	Pro	Ile
35					245					250						255
	Pro	Asp	Pro	Ile	Leu	Glu	Arg	Lys	Lys	Val	Leu	Lys	Val	Tyr	Asp	Pro
					260				265					270		
40	Ser	Gln	His	Asp	Tyr	Glu	Thr	Asp	Lys	Pro	Ser	Met	Val	Glu	Ile	Arg
				275				280						285		
	Pro	Gly	His	Tyr	Val	Trp	Ala	Asn	Gln	Ala	Glu	Leu	Ala			
45				290			295					300				

(2) INFORMATION FOR SEQ ID NO:163:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 151 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

WO 98/26072

PCT/US97/22578

-250-

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

5 Gln Ile Gln Lys Ser Phe Lys Gly Gln Ser Pro Tyr Gly Lys Leu Tyr  
1 5 10

Leu Val Ala Thr Pro Ile Gly Asn Leu Asp Asp Met Thr Phe Arg Ala  
20 25 30

10 Ile Gln Thr Leu Lys Glu Val Asp Trp Ile Ala Ala Glu Asp Thr Arg  
35 40 45

Asn Thr Gly Leu Leu Leu Lys His Phe Asp Ile Ser Thr Lys Gln Ile  
50 55 60

15 Ser Phe His Glu His Asn Ala Lys Glu Lys Ile Pro Asp Leu Ile Gly  
65 70 75 80

20 Phe Leu Lys Ala Gly Gln Ser Ile Ala Gln Val Ser Asp Ala Gly Leu  
85 90 95

Pro Ser Ile Ser Asp Pro Gly His Asp Leu Val Lys Ala Ala Ile Glu  
100 105 110

25 Glu Glu Ile Ala Val Val Thr Val Pro Gly Ala Ser Ala Gly Ile Ser  
115 120 125

Ala Leu Ile Ala Ser Gly Leu Ala Pro Gln Pro His Ile Phe Tyr Gly  
130 135 140

30 Phe Leu Pro Arg Lys Ser Gly  
145 150

## (2) INFORMATION FOR SEQ ID NO:164:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 258 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:

55 Ser Arg Lys Asp Lys Gln Glu Arg Ile Ser Lys Glu Thr Met Glu Ile  
1 5 10 15

Tyr Ala Pro Leu Ala His Arg Leu Gly Ile Ser Ser Val Lys Trp Glu  
20 25 30

60 Leu Glu Asp Leu Ser Phe Arg Tyr Leu Asn Pro Thr Glu Phe Tyr Lys  
35 40 45



WO 98/26072

PCT/US97/22578

-251-

Ile Thr His Met Met Lys Glu Lys Arg Arg Glu Arg Glu Ala Leu Val  
 50 55 60  
 5 Asp Glu Val Val Thr Lys Leu Glu Glu Tyr Thr Thr Glu Arg His Leu  
 65 70 75 80  
 Lys Gly Lys Ile Tyr Gly Arg Pro Lys His Ile Tyr Ser Ile Phe Arg  
 85 90 95  
 10 Lys Met Gln Asp Lys Arg Lys Arg Phe Glu Glu Ile Tyr Asp Leu Ile  
 100 105 110  
 Ala Ile Arg Cys Ile Leu Asp Thr Gln Ser Asp Val Tyr Ala Met Leu  
 115 120 125  
 15 Gly Tyr Val His Glu Phe Trp Lys Pro Met Pro Gly Arg Phe Lys Asp  
 130 135 140  
 Tyr Ile Ala Asn Arg Lys Ala Asn Gly Tyr Gln Ser Ile His Thr Thr  
 145 150 155 160  
 Val Tyr Gly Pro Lys Gly Pro Ile Glu Phe Gln Ile Arg Thr Lys Glu  
 165 170 175  
 25 Met His Glu Val Ala Glu Tyr Gly Val Ala Ala His Trp Ala Tyr Lys  
 180 185 190  
 Lys Gly Ile Lys Gly Gln Val Asn Ser Lys Glu Ser Ala Ile Gly Met  
 195 200 205  
 30 Asn Trp Ile Lys Glu Met Met Glu Leu Gln Asp Gln Ala Asp Asp Ala  
 210 215 220  
 35 Lys Glu Phe Val Asp Ser Val Lys Glu Asn Tyr Leu Ala Glu Glu Ile  
 225 230 235 240  
 Thr Val Leu Pro Gln Met Glu Leu Ser Val Pro Ser Gln Arg Phe Arg  
 245 250 255  
 40 Thr Asp

## (2) INFORMATION FOR SEQ ID NO:165:

- 45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 289 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant  
 50  
 (ii) MOLECULE TYPE: peptide  
 (iii) HYPOTHETICAL: NO  
 55 (iv) ANTI-SENSE: NO  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:  
 60 Thr Lys Val Gly Gly Glu Ala Asp Tyr Leu Val Phe Pro Arg Asn Arg  
 1 5 10 15

WO 98/26072

PCT/US97/22578

-252-

Phe Glu Leu Ala Arg Val Val Lys Phe Ala Asn Gln Glu Asn Ile Pro  
 20 25 30  
 5 Trp Met Val Leu Gly Asn Ala Ser Asn Ile Ile Val Arg Asp Gly Gly  
 35 40 45  
 Ile Arg Gly Phe Val Ile Leu Cys Asp Lys Leu Asn Val Ser Val  
 50 55 60  
 10 Asp Gly Tyr Thr Ile Glu Ala Glu Ala Gly Ala Asn Leu Ile Glu Thr  
 65 70 75 80  
 15 Thr Arg Ile Ala Leu Arg His Ser Leu Thr Gly Phe Glu Phe Ala Cys  
 85 90 95  
 Gly Ile Pro Gly Ser Val Gly Gly Ala Val Phe Met Asn Ala Gly Ala  
 100 105 110  
 20 Tyr Gly Gly Glu Ile Ala His Ile Leu Gln Ser Cys Lys Val Leu Thr  
 115 120 125  
 Lys Asp Gly Glu Ile Glu Thr Leu Ser Ala Lys Asp Leu Ala Phe Gly  
 130 135 140  
 25 Tyr Arg His Ser Ala Ile Gln Glu Ser Gly Ala Val Val Leu Ser Val  
 145 150 155 160  
 30 Lys Phe Ala Leu Ala Pro Gly Thr His Gln Val Ile Lys Gln Glu Met  
 165 170 175  
 Asp Arg Leu Thr His Leu Arg Glu Leu Lys Gln Pro Leu Glu Tyr Pro  
 180 185 190  
 35 Ser Cys Gly Ser Val Phe Lys Arg Pro Val Gly His Phe Ala Gly Gln  
 195 200 205  
 Leu Ile Ser Glu Ala Gly Leu Lys Gly Tyr Arg Ile Gly Gly Val Glu  
 210 215 220  
 40 Val Ser Glu Lys His Ala Gly Phe Met Ile Asn Val Ala Asp Gly Thr  
 225 230 235 240  
 45 Ala Lys Asp Tyr Glu Asp Leu Ile Gln Ser Val Ile Glu Lys Val Lys  
 245 250 255  
 Glu His Ser Gly Ile Thr Leu Glu Arg Glu Val Arg Ile Leu Gly Glu  
 260 265 270  
 50 Ser Leu Ser Val Ala Lys Met Tyr Ala Gly Gly Phe Thr Pro Cys Lys  
 275 280 285

Arg

55

(2) INFORMATION FOR SEQ ID NO:166:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 114 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: not relevant

60

WO 98/26072

PCT/US97/22578

-253-

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

5 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

Ala Lys Arg Arg Lys Leu Val Lys Ser Thr Thr Leu Leu Leu Ala Cys  
1 5 10 15  
Leu Gln Lys Pro Phe Leu Thr Thr Leu Leu Pro Thr Ile Trp Ile Cys  
20 25 30  
Val Lys Ser Ser Met Phe Thr Leu Arg Leu Asn Thr Trp Ile Lys  
35 40 45  
Asp Phe His Ser Pro Ser Ser Cys Val Thr Phe Gln Lys Ala Phe  
50 55 60  
Thr Asn Gly Arg Gly Lys Ile Asn Lys Arg His Val Thr Cys Pro Ser  
65 70 75 80  
Phe Val Thr Met Pro Leu Thr Arg Glu Ser Ser Leu Ser Thr Thr Ser  
85 90 95  
Val Pro Leu Gln Met Thr Val Glu Lys Ser Ala Pro Thr Asn Val Lys  
100 105 110  
Ala Val

35 (2) INFORMATION FOR SEQ ID NO:167:

(i) SEQUENCE CHARACTERISTICS:  
40 (A) LENGTH: 253 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

45 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:

Met Leu Lys Gln Glu Lys Leu Ala Lys Ile Leu Glu Ile Val Asn Ser  
1 5 10 15  
Lys Gly Thr Ile Thr Val Lys Gln Ile Met Asp Glu Ile Ala Val Ser  
20 25 30  
Asp Met Thr Ala Arg Arg Tyr Leu Gln Glu Leu Ala Asp Lys Asp Leu  
35 40 45

60

WO 98/26072

PCT/US97/22578

-254-

Leu Ile Arg Val His Gly Gly Ala Glu Lys Leu Arg Thr Asn Ser Leu  
50 55 60

5 Leu Thr Asn Glu Arg Ser Asn Ile Glu Lys Gln Ala Leu Gln Thr Ala  
65 70 75 80

Glu Lys Gln Glu Ile Ala His Phe Ala Gly Ser Leu Val Glu Glu Arg  
85 90 95

10 Glu Thr Ile Phe Ile Gly Pro Gly Thr Thr Leu Glu Phe Phe Ala Arg  
100 105 110

Glu Leu Pro Ile Asp Asn Ile Arg Val Val Thr Asn Ser Leu Pro Val  
115 120 125

15 Phe Leu Ile Leu Ser Glu Arg Lys Leu Thr Asp Leu Ile Leu Ile Gly  
130 135 140

Gly Asn Tyr Arg Asp Ile Thr Gly Ala Phe Val Gly Thr Leu Thr Leu  
145 150 155 160

Gln Asn Leu Ser Asn Leu Gln Phe Ser Lys Ala Phe Val Ser Cys Asn  
165 170 175

25 Gly Ile Gln Asn Gly Ala Leu Ala Thr Phe Ser Glu Glu Glu Gly Glu  
180 185 190

Ala Gln Arg Ile Ala Leu Asn Asn Ser Asn Lys Lys Tyr Leu Leu Ala  
195 200 205

30 Asp His Ser Lys Phe Asn Lys Phe Asp Phe Tyr Thr Phe Tyr Asn Ile  
210 215 220

35 Ser Asn Leu Asp Thr Ile Val Ser Asp Ser Lys Leu Ser Asp Ser Ile  
225 230 235 240

Leu Phe Lys Leu Ser Lys His Ile Lys Val Ile Lys Pro  
245 250

- 40 (2) INFORMATION FOR SEQ ID NO:168:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 320 amino acids
- (B) TYPE: amino acid
- 45 (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide
- 50 (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- 55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

Met Glu Thr Tyr Tyr Lys Ala Ile Asn Trp Asn Ala Ile Glu Asp Val  
1 5 10 15

60 Ile Asp Lys Ser Thr Trp Glu Lys Leu Thr Glu Gln Phe Trp Leu Asp  
20 25 30

WO 98/26072

PCT/US97/22578

-255-

	Thr	Arg	Ile	Pro	Leu	Ser	Asn	Asp	Leu	Asp	Asp	Trp	Arg	Lys	Leu	Ser
					35				40					45		
5	Asn	Lys	Glu	Lys	Asp	Leu	Val	Gly	Lys	Val	Phe	Gly	Gly	Leu	Thr	Leu
		50					55					60				
	Leu	Asp	Thr	Met	Gln	Ser	Glu	Thr	Gly	Val	Gln	Ala	Leu	Arg	Ala	Asp
		65				70					75				80	
10	Ile	Arg	Thr	Pro	His	Glu	Glu	Ala	Val	Phe	Asn	Asn	Ile	Gln	Phe	Met
					85					90					95	
	Glu	Ser	Val	His	Ala	Lys	Ser	Tyr	Ser	Ser	Ile	Phe	Ser	Thr	Leu	Asn
				100					105					110		
15	Thr	Lys	Ala	Glu	Ile	Glu	Glu	Ile	Phe	Glu	Trp	Thr	Asn	Thr	Asn	Pro
			115					120					125			
20	Tyr	Leu	Gln	Lys	Lys	Ala	Glu	Ile	Val	Asn	Glu	Ile	Tyr	Leu	Asn	Gly
		130					135					140				
	Ser	Pro	Leu	Glu	Lys	Lys	Val	Ala	Ser	Val	Phe	Leu	Glu	Thr	Phe	Ileu
		145				150				155					160	
25	Phe	Tyr	Ser	Gly	Phe	Phe	Thr	Pro	Leu	Tyr	Tyr	Leu	Gly	Asn	Asn	Lys
				165						170				175		
30	Leu	Ala	Asn	Val	Ala	Glu	Ile	Ile	Lys	Leu	Ile	Ile	Arg	Asp	Glu	Ser
				180				185						190		
	Val	His	Gly	Thr	Tyr	Ile	Gly	Tyr	Lys	Phe	Gln	Leu	Gly	Phe	Asn	Glu
			195				200						205			
35	Leu	Pro	Glu	Glu	Glu	Gln	Glu	Lys	Leu	Lys	Glu	Trp	Met	Tyr	Asp	Leu
			210				215					220				
	Leu	Tyr	Thr	Leu	Tyr	Glu	Asn	Glu	Glu	Gly	Tyr	Thr	Glu	Ser	Leu	Tyr
		225				230				235					240	
40	Asp	Gly	Val	Gly	Trp	Thr	Glu	Glu	Val	Lys	Thr	Phe	Leu	Arg	Tyr	Asn
				245					250					255		
	Ala	Asn	Lys	Ala	Leu	Met	Asn	Met	Gly	Gln	Asp	Pro	Leu	Phe	Pro	Asp
				260				265					270			
	Ser	Ala	Glu	Asp	Val	Asn	Pro	Ile	Val	Met	Asn	Gly	Ile	Ser	Thr	Gly
			275				280					285				
50	Thr	Ser	Asn	His	Asp	Phe	Phe	Ser	Gln	Val	Gly	Asn	Gly	Tyr	Leu	Leu
			290				295					300				
	Gly	Glu	Val	Glu	Ala	Met	Gln	Asp	Asp	Asp	Tyr	Asn	Tyr	Gly	Leu	Asp
		305				310					315				320	
55																

(2) INFORMATION FOR SEQ ID NO:169:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 240 amino acids

(B) TYPE: amino acid

60

WO 98/26072

PCT/US97/22578

-256-

(C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:

Ile	Glu	Glu	Gly	Val	Lys	Val	Val	Thr	Thr	Gly	Ala	Gly	Asn	Pro	Ser
1			5					10					15		

Lys	Tyr	Met	Glu	Arg	Phe	His	Glu	Ala	Gly	Ile	Ile	Val	Ile	Pro	Val
	20						25					30			

Val	Pro	Ser	Val	Ala	Leu	Ala	Lys	Arg	Met	Glu	Lys	Ile	Gly	Ala	Asp
	35						40				45				

Ala	Val	Ile	Ala	Glu	Gly	Met	Glu	Ala	Gly	Gly	His	Ile	Gly	Lys	Leu
	50					55					60				

Thr	Thr	Met	Thr	Leu	Val	Arg	Gln	Val	Ala	Thr	Ala	Val	Ser	Ile	Pro
	65			70						75				80	

Val	Ile	Ala	Ala	Gly	Gly	Ile	Ala	Asp	Gly	Glu	Gly	Ala	Ala	Gly	
		85							90				95		

Phe	Met	Leu	Gly	Ala	Glu	Ala	Val	Gln	Val	Gly	Thr	Arg	Phe	Val	Val
		100						105					110		

Ala	Lys	Glu	Ser	Asn	Ala	His	Pro	Asn	Tyr	Lys	Glu	Lys	Ile	Leu	Lys
	115						120				125				

Ala	Arg	Asp	Ile	Asp	Thr	Thr	Ile	Ser	Ala	Gln	His	Phe	Gly	His	Ala
	130				135						140				

Val	Arg	Ala	Ile	Lys	Asn	Gln	Leu	Thr	Arg	Asp	Phe	Glu	Leu	Ala	Glu
	145			150					155				160		

Lys	Asp	Ala	Phe	Lys	Gln	Glu	Asp	Pro	Asp	Leu	Glu	Ile	Phe	Glu	Gln
		165						170				175			

Met	Gly	Ala	Gly	Ala	Leu	Ala	Lys	Ala	Val	Val	His	Gly	Asp	Val	Glu
		180					185					190			

Gly	Gly	Ser	Val	Met	Ala	Gly	Gln	Ile	Ala	Gly	Leu	Val	Ser	Lys	Glu
	195				200						205				

Glu	Thr	Ala	Glu	Glu	Ile	Leu	Lys	Asp	Leu	Tyr	Tyr	Gly	Ala	Ala	Lys
	210				215					220					

Lys	Ile	Gln	Glu	Glu	Ala	Ser	Arg	Trp	Thr	Gly	Val	Val	Arg	Asn	Asp
	225				230					235				240	

(2) INFORMATION FOR SEQ ID NO:170:

(i) SEQUENCE CHARACTERISTICS:

WO 98/26072

PCT/US97/22578

-257-

(A) LENGTH: 243 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

5 (ii) MOLECULE TYPE: peptide  
 (iii) HYPOTHETICAL: NO  
 10 (iv) ANTI-SENSE: NO  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:  
 15 Met Lys Leu Glu His Lys Asn Ile Phe Ile Thr Gly Ser Ser Arg Gly  
     1                    5                    10                    15  
     Ile Gly Leu Ala Ile Ala His Lys Phe Ala Gln Ala Gly Ala Asn Ile  
                     20                    25                    30  
 20 Val Leu Asn Ser Arg Gly Ala Ile Ser Glu Glu Leu Leu Ala Glu Phe  
     35                    40                    45  
     Ser Asn Tyr Gly Ile Lys Val Val Pro Ile Ser Gly Asp Val Ser Asp  
     50                    55                    60  
     Phe Ala Asp Ala Lys Arg Met Ile Asp Gln Ala Ile Ala Glu Leu Gly  
     65                    70                    75                    80  
 30 Ser Val Asp Val Leu Val Asn Asn Ala Gly Ile Thr Gln Asp Thr Leu  
     85                    90                    95  
     Met Leu Lys Met Thr Glu Ala Asp Phe Glu Lys Val Leu Lys Val Asn  
     100                    105                    110  
     Leu Thr Gly Ala Phe Asn Met Thr Gln Ser Val Leu Lys Pro Met Met  
     115                    120                    125  
 40 Lys Ala Arg Glu Gly Ala Ile Ile Asn Met Ser Ser Val Val Gly Leu  
     130                    135                    140  
     Met Gly Asn Ile Gly Gln Ala Asn Tyr Ala Ala Ser Lys Ala Gly Leu  
     145                    150                    155                    160  
 45 Ile Gly Phe Thr Lys Ser Val Ala Arg Glu Val Ala Ser Arg Asn Ile  
     165                    170                    175  
     Arg Val Asn Val Ile Ala Pro Gly Met Ile Glu Ser Asp Met Thr Ala  
     180                    185                    190  
 50 Ile Leu Ser Asp Lys Ile Lys Glu Ala Thr Leu Ala Gln Ile Pro Met  
     195                    200                    205  
 55 Lys Glu Phe Gly Gln Ala Glu Gln Val Ala Asp Leu Thr Val Phe Leu  
     210                    215                    220  
     Ala Gly Gln Asp Tyr Leu Thr Gly Gln Val Ile Ala Ile Asp Gly Gly  
     225                    230                    235                    240  
 60 Leu Ser Met

WO 98/26072

PCT/US97/22578

-258-

## (2) INFORMATION FOR SEQ ID NO:171:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 306 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

Met	Thr	Lys	Thr	Ala	Phe	Leu	Phe	Ala	Gly	Gln	Gly	Ala	Gln	Tyr	Leu	1	5	10	15
Gly	Met	Gly	Arg	Asp	Phe	Tyr	Asp	Gln	Tyr	Pro	Ile	Val	Lys	Glu	Thr	20	25	30	
Ile	Asp	Arg	Ala	Ser	Gln	Val	Leu	Gly	Tyr	Asp	Leu	Arg	Tyr	Leu	Ile	35	40	45	
Asp	Thr	Glu	Glu	Asp	Lys	Leu	Asn	Gln	Thr	Arg	Tyr	Thr	Gln	Pro	Ala	50	55	60	
Ile	Leu	Ala	Thr	Ser	Val	Ala	Ile	Tyr	Arg	Leu	Leu	Gln	Glu	Lys	Gly	65	70	75	80
Tyr	Gln	Pro	Asp	Met	Val	Ala	Gly	Leu	Ser	Leu	Gly	Glu	Tyr	Ser	Ala	85	90	95	
Leu	Val	Ala	Ser	Gly	Ala	Leu	Asp	Phe	Glu	Asp	Ala	Val	Ala	Leu	Val	100	105	110	
Ala	Lys	Arg	Gly	Ala	Tyr	Met	Glu	Glu	Ala	Ala	Pro	Ala	Asp	Ser	Gly	115	120	125	
Lys	Met	Val	Ala	Val	Leu	Asn	Thr	Pro	Val	Glu	Val	Ile	Glu	Glu	Ala	130	135	140	
Cys	Gln	Lys	Ala	Ser	Glu	Leu	Gly	Val	Val	Thr	Pro	Ala	Asn	Tyr	Asn	145	150	155	160
Thr	Pro	Ala	Gln	Ile	Val	Ile	Ala	Gly	Glu	Val	Val	Ala	Val	Asp	Arg	165	170	175	
Ala	Val	Glu	Leu	Leu	Gln	Glu	Ala	Gly	Ala	Lys	Arg	Leu	Ile	Pro	Leu	180	185	190	
Lys	Val	Ser	Gly	Pro	Phe	His	Thr	Ser	Leu	Leu	Glu	Pro	Ala	Ser	Gln	195	200	205	
Lys	Leu	Ala	Glu	Thr	Leu	Ala	Gln	Val	Ser	Phe	Ser	Asp	Phe	Thr	Cys	210	215	220	
Pro	Leu	Val	Gly	Asn	Thr	Glu	Ala	Ala	Val	Met	Gln	Lys	Glu	Asp	Ile	225	230	235	240



-259-

	225		230		235		240
	Ala	Gln	Leu	Leu	Thr	Arg	Gln
				245			250
5					Glu	Pro	Val
					Arg	Phe	Tyr
						255	Glu
	Ser	Ile	Gly	Val	Met	Gln	Glu
				260			265
					Ile	Ser	Asn
					Phe	Ile	Glu
						270	Ile
10	Gly	Pro	Gly	Lys	Val	Leu	Ser
				275			280
					Gly	Phe	Val
					Lys	Lys	Ile
						285	Asp
					Gln	Thr	
	Ala	His	Leu	Ala	His	Val	Glu
				290			295
					Asp	Gln	Ala
					Ser	Leu	Val
						300	Ala
							Leu
							Leu
15	Glu	Lys					
	305						

(2) INFORMATION FOR SEQ ID NO:172:

20	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 318 amino acids
	(B) TYPE: amino acid
	(C) STRANDEDNESS: not relevant
	(D) TOPOLOGY: not relevant
25	(ii) MOLECULE TYPE: peptide
	(iii) HYPOTHETICAL: NO
30	(iv) ANTI-SENSE: NO
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:
35	Met Lys Leu Asn Arg Val Val Val Thr Gly Tyr Gly Val Thr Ser Pro
	1 5 10 15
	Ile Gly Asn Thr Pro Glu Glu Phe Trp Asn Ser Leu Ala Thr Gly Lys
40	20 25 30
	Ile Gly Ile Gly Gly Ile Thr Lys Phe Asp His Ser Asp Phe Asp Val
	35 40 45
45	His Asn Ala Ala Glu Ile Gln Asp Phe Pro Phe Asp Lys Tyr Phe Val
	50 55 60
	Lys Lys Asp Thr Asn Arg Phe Asp Asn Tyr Ser Leu Tyr Ala Leu Tyr
	65 70 75 80
50	Ala Ala Gln Glu Ala Val Asn His Ala Asn Leu Asp Val Glu Ala Leu
	85 90 95
	Asn Arg Asp Arg Phe Gly Val Ile Val Ala Ser Gly Ile Gly Gly Ile
55	100 105 110
	Lys Glu Ile Glu Asp Gln Val Leu Arg Leu His Glu Lys Gly Pro Lys
	115 120 125
60	Arg Val Lys Pro Met Thr Leu Pro Lys Ala Leu Pro Asn Met Ala Ser
	130 135 140

WO 98/26072

PCT/US97/22578

-260-

Gly Asn Val Ala Met Arg Phe Gly Ala Asn Gly Val Cys Lys Ser Ile  
 145 150 155 160  
 5 Asn Thr Ala Cys Ser Ser Ser Asn Asp Ala Ile Gly Asp Ala Phe Arg  
 165 170 175  
 Ser Ile Lys Phe Gly Phe Gln Asp Val Met Leu Val Gly Gly Thr Glu  
 180 185 190  
 10 Ala Ser Ile Thr Pro Phe Ala Ile Ala Gly Phe Gln Ala Leu Thr Ala  
 195 200 205  
 Leu Ser Thr Thr Glu Asp Pro Thr Arg Ala Ser Ile Pro Phe Asp Lys  
 210 215 220  
 15 Asp Arg Asn Gly Phe Val Met Gly Glu Gly Ser Gly Met Leu Val Leu  
 225 230 235 240  
 20 Glu Ser Leu Glu His Ala Glu Lys Arg Gly Ala Thr Ile Leu Ala Glu  
 245 250 255  
 Val Val Gly Tyr Gly Asn Thr Cys Asp Ala Tyr His Met Thr Ser Pro  
 260 265 270  
 25 His Pro Glu Gly Gln Gly Ala Ile Lys Ala Ile Lys Leu Ala Leu Glu  
 275 280 285  
 Glu Ala Glu Ile Ser Pro Glu Gln Val Ala Met Leu Met Leu Thr Glu  
 290 295 300  
 30 Arg Gln Leu Leu Pro Met Lys Lys Glu Lys Val Val Leu Ser  
 305 310 315

## (2) INFORMATION FOR SEQ ID NO:173:

- 35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 396 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 40 (D) TOPOLOGY: not relevant  
 (ii) MOLECULE TYPE: peptide  
 (iii) HYPOTHETICAL: NO  
 45 (iv) ANTI-SENSE: NO  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:  
 50 Met Gln Ala Val Glu His Phe Ile Lys Gln Phe Val Pro Glu His Tyr  
 1 5 10 15  
 55 Asp Leu Phe Leu Asp Leu Ser Arg Glu Thr Lys Thr Phe Ser Gly Lys  
 20 25 30  
 Val Thr Ile Thr Gly Gln Ala Gln Ser Asp Arg Ile Ser Leu His Gln  
 35 40 45  
 60 Lys Asp Leu Glu Ile Thr Ser Val Glu Val Ala Gly Gln Ala Arg Pro  
 50 55 60

WO 98/26072

PCT/US97/22578

-261-

	Phe Thr Val Asp His Asp Asn Glu Ala Leu His Ile Glu Leu Ala Glu	65	76	75	80
5	Ala Gly Gln Val Glu Leu Val Leu Ala Phe Ser Gly Lys Ile Thr Asp	85	90	95	
	Asn Met Thr Gly Ile Tyr Pro Ser Tyr Tyr Thr Val Asp Gly Val Lys	100	105	110	
10	Lys Glu Val Leu Ser Thr Gln Phe Glu Ser His Phe Ala Arg Glu Ala	115	120	125	
	Phe Pro Cys Val Asp Glu Pro Glu Ala Lys Ala Thr Phe Asp Leu Ser	130	135	140	
15	Leu Arg Phe Asp Gln Ala Glu Gly Glu Leu Ala Leu Ser Asn Met Pro	145	150	155	
	Glu Ile Asp Val Glu Asn Arg Lys Glu Thr Gly Ile Trp Lys Phe Glu	165	170	175	
20	Thr Thr Pro Arg Met Ser Ser Tyr Leu Leu Ala Phe Val Ala Gly Asp	180	185	190	
25	Leu Gln Gly Val Thr Ala Lys Thr Lys Asn Gly Thr Leu Val Gly Val	195	200	205	
	Tyr Ser Thr Lys Ala His Pro Leu Ser Asn Leu Asp Phe Ser Leu Asp	210	215	220	
30	Ile Ala Val Arg Ser Ile Glu Phe Tyr Glu Asp Tyr Tyr Gly Val Lys	225	230	235	240
	Tyr Pro Ile Pro Gln Ser Leu His Ile Ala Leu Pro Asp Phe Ser Ala	245	250	255	
35	Gly Ala Met Glu Asn Trp Gly Leu Val Thr Tyr Arg Glu Val Tyr Leu	260	265	270	
40	Val Val Asp Glu Asn Ser Thr Phe Ala Ser Arg Gln Gln Val Ala Leu	275	280	285	
	Val Val Ala His Glu Leu Ala His Gln Trp Phe Gly Asn Leu Val Thr	290	295	300	
45	Met Lys Trp Trp Asp Asn Leu Trp Leu Asn Glu Ser Phe Ala Asn Met	305	310	315	320
	Met Glu Tyr Val Cys Val Asp Thr Ile Glu Pro Ser Trp Asn Ile Phe	325	330	335	
50	Glu Asp Phe Gln Thr Gly Gly Val Pro Leu Ala Leu Glu Arg Asp Ala	340	345	350	
55	Thr Asp Gly Val Gln Ser Val His Val Glu Val Lys His Pro Asp Glu	355	360	365	
	Ile Asn Thr Leu Phe Asp Gly Ala Ile Val Tyr Ala Arg Lys Arg Leu	370	375	380	
60					

-262-

Met His Met Leu Arg Val Ala Arg Asp Ala Asp Leu  
 385 390 395

## (2) INFORMATION FOR SEQ ID NO:174:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 127 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

Met Asp Phe Leu Leu Phe Tyr Asp Ser Lys Lys Lys Gly Asp Thr Met  
 1 5 10 15

Thr Tyr Leu Glu Lys Trp Phe Asp Phe Asn Arg Arg Gln Lys Glu Ile  
 20 25 30

Glu Ser Leu Leu Glu Glu Thr Ile Ala Gln Gln Ser Glu Gln Ser Leu  
 35 40 45

Thr Leu Lys Glu Phe Tyr Leu Leu Tyr Tyr Leu Asp Leu Ala Glu Glu  
 50 55 60

Lys Ser Leu Arg Gln Ile Asp Leu Pro Asp Lys Leu His Leu Ser Pro  
 65 70 75 80

Ser Ala Val Ser Arg Met Val Ala Arg Leu Glu Ala Lys Asn Cys Gly  
 85 90 95

Leu Leu Ser Arg Met Cys Cys His Gln Asp Arg Arg Ser Ser Phe Ile  
 100 105 110

Cys Leu Thr Asn Asp Gly Gln Lys Thr Leu Ala Ser Leu Gln Lys  
 115 120 125

## (2) INFORMATION FOR SEQ ID NO:175:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 373 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

WO 98/26072

PCT/US97/22578

-263-

	Met	Leu	Tyr	Asp	Tyr	Gly	Asn	Ser	Val	Trp	Leu	Ala	Ser	Met	Gly	Thr	
	1			5						10					15		
5	Ile	Gly	Gln	Thr	Val	Leu	Gly	Met	Tyr	Gln	Ile	Ser	Glu	Leu	Val	Thr	
				20					25					30			
	Ser	Ile	Leu	Val	Asn	Pro	Phe	Gly	Gly	Val	Ile	Ser	Asp	Arg	Phe	Ser	
			35					40					45				
10	Arg	Arg	Lys	Ile	Leu	Met	Thr	Ala	Asp	Leu	Val	Cys	Gly	Ile	Leu	Cys	
		50					55					60					
	Leu	Ala	Ile	Ser	Phe	Ile	Arg	Asn	Asp	Ser	Trp	Met	Ile	Gly	Ala	Leu	
		65				70				75					80		
15	Ile	Val	Ala	Asn	Ile	Val	Gln	Ala	Ile	Ala	Phe	Ala	Phe	Ser	Arg	Thr	
				85					90					95			
	Ala	Asn	Lys	Ala	Ile	Ile	Thr	Glu	Val	Val	Glu	Lys	Asn	Glu	Ile	Val	
			100					105					110				
20	Ile	Tyr	Asn	Ser	Arg	Leu	Glu	Leu	Val	Leu	Gln	Val	Val	Gly	Val	Ser	
			115				120						125				
25	Ser	Pro	Val	Leu	Ser	Phe	Leu	Val	Leu	Gln	Phe	Ala	Ser	Leu	His	Met	
			130				135					140					
	Thr	Leu	Leu	Leu	Asp	Ser	Leu	Thr	Phe	Phe	Ile	Ala	Phe	Val	Leu	Val	
					150						155				160		
30	Ala	Phe	Leu	Pro	Lys	Glu	Glu	Ala	Lys	Val	Gln	Glu	Lys	Lys	Ala	Phe	
				165					170					175			
	Thr	Gly	Arg	Asp	Ile	Phe	Val	Asp	Ile	Lys	Asp	Gly	Leu	His	Tyr	Ile	
			180						185				190				
35	Trp	His	Gln	Gln	Glu	Ile	Phe	Phe	Leu	Leu	Leu	Val	Ala	Ser	Ser	Val	
			195				200					205					
40	Asn	Phe	Phe	Phe	Ala	Ala	Phe	Glu	Phe	Leu	Leu	Pro	Phe	Ser	Asn	Gln	
			210				215					220					
	Leu	Tyr	Gly	Ser	Glu	Gly	Ala	Tyr	Ala	Ser	Ile	Leu	Thr	Met	Gly	Ala	
			225			230					235				240		
45	Ile	Gly	Ser	Ile	Ile	Gly	Ala	Leu	Leu	Ala	Ser	Lys	Ile	Lys	Ala	Asn	
				245						250				255			
	Ile	Tyr	Asn	Leu	Leu	Ile	Leu	Leu	Ala	Leu	Thr	Gly	Val	Gly	Val	Phe	
			260				265						270				
50	Met	Met	Gly	Leu	Pro	Leu	Pro	Thr	Phe	Leu	Ser	Phe	Ser	Gly	Asn	Leu	
			275				280					285					
55	Val	Cys	Glu	Leu	Phe	Met	Thr	Ile	Phe	Asn	Ile	His	Phe	Phe	Thr	Gln	
			290				295				300						
	Val	Gln	Thr	Lys	Val	Glu	Ser	Glu	Phe	Leu	Gly	Arg	Val	Leu	Ser	Thr	
			305			310					315				320		
60	Ile	Phe	Thr	Leu	Ala	Ile	Leu	Phe	Met	Pro	Ile	Ala	Lys	Gly	Phe	Met	

WO 98/26072

PCT/US97/22578

-264-

325 330 335

Thr Val Leu Pro Ser Val His Leu Ser Ser Phe Leu Ile Ile Gly Ser  
340 345 350

5 Gly Val Ile Ile Leu Ser Cys Ile Ser Phe Ile Tyr Val Arg Thr His  
355 360 365

Phe Glu Lys Leu Ile  
370

10 (2) INFORMATION FOR SEQ ID NO:176:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 427 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

15 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

30 Met Ser Val Ser Phe Glu Asn Lys Glu Thr Asn Arg Gly Val Leu Thr  
1 5 10 15

Phe Thr Ile Ser Gln Asp Gln Ile Lys Pro Glu Leu Asp Arg Val Phe  
35 20 25 30

Lys Ser Val Lys Lys Ser Leu Asn Val Pro Gly Phe Arg Lys Gly His  
35 40 45

40 Leu Pro Arg Pro Ile Phe Asp Gln Lys Phe Gly Glu Glu Ala Leu Tyr  
50 55 60

Gln Asp Ala Met Asn Ala Leu Leu Pro Asn Ala Tyr Glu Ala Ala Val  
65 70 75 80

45 Lys Glu Ala Gly Leu Glu Val Val Ala Gln Pro Lys Ile Asp Val Thr  
85 90 95

50 Ser Met Glu Lys Gly Gln Asp Trp Val Ile Thr Ala Glu Val Val Thr  
100 105 110

Lys Pro Glu Val Lys Leu Gly Asp Tyr Lys Asn Leu Glu Val Ser Val  
115 120 125

55 Asp Val Glu Lys Glu Val Thr Asp Ala Asp Val Glu Glu Arg Ile Glu  
130 135 140

Arg Glu Arg Asn Asn Leu Ala Glu Leu Val Ile Lys Glu Ala Ala Ala  
145 150 155 160

60 Glu Asn Gly Asp Thr Val Val Ile Asp Phe Val Gly Ser Ile Asp Gly

WO 98/26072

PCT/US97/22578

-265-

	165	170	175
	Val Glu Phe Asp Gly Gly Lys Gly	Glu Asn Phe Ser Leu Gly Leu Gly	
	180	185	190
5	Ser Gly Gln Phe Ile Pro Gly Phe	Glu Asp Gln Leu Val Gly His Ser	
	195	200	205
	Ala Gly Glu Thr Val Asp Val Ile Val Thr Phe	Pro Glu Asp Tyr Gln	
10	210	215	220
	Ala Glu Asp Leu Ala Gly Lys Glu Ala Lys Phe	Val Thr Thr Ile His	
	225	230	240
15	Glu Val Lys Ala Lys Glu Val Pro Ala Leu Asp Asp	Glu Leu Ala Lys	
	245	250	255
	Asp Ile Asp Glu Glu Val Glu Thr Leu	Ala Asp Leu Lys Glu Lys Tyr	
20	260	265	270
	Arg Lys Glu Leu Ala Ala Ala Lys Glu	Glu Thr Tyr Lys Asp Ala Val	
	275	280	285
25	Glu Gly Ala Ala Ile Asp Thr Ala Val	Glu Asn Ala Glu Ile Val Glu	
	290	295	300
	Leu Pro Glu Glu Met Ile His Glu Glu Val His	Arg Ser Val Asn Glu	
	305	310	315
30	Phe Leu Gly Asn Leu Gln Arg Gln Gly Ile	Asn Pro Asp Met Tyr Phe	
	325	330	335
	Gln Ile Thr Gly Thr Thr Gln Glu Asp	Leu His Asn Gln Tyr Gln Ala	
35	340	345	350
	Glu Ala Glu Ser Arg Thr Lys Thr	Asn Leu Val Ile Glu Ala Val Ala	
	355	360	365
40	Lys Ala Glu Gly Phe Asp Ala Ser Glu Glu Glu	Ile Gln Lys Glu Val	
	370	375	380
	Glu Gln Leu Ala Ala Asp Tyr Asn Met Glu	Val Ala Gln Val Gln Asn	
	385	390	395
45	Leu Leu Ser Ala Asp Met Leu Lys His Asp	Ile Thr Ile Lys Lys Ala	
	405	410	415
	Val Glu Leu Ile Thr Ser Thr Ala Thr	Val Lys	
50	420	425	

(2) INFORMATION FOR SEQ ID NO:177:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 207 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

WO 98/26072

PCT/US97/22578

-266-

(iv) ANTI-SENSE: NO

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:  
 Gly Gly Asp Lys Asp Phe Leu Thr Ser Ile Cys Leu Thr Asn Asp Pro  
 1 5 10 15  
 10 Phe Leu Gly Phe Arg Ala Leu Arg Ile Ser Ile Ser Glu Thr Gly Asp  
 20 25 30  
 15 Ala Met Phe Arg Thr Gln Ile Arg Ala Leu Leu Arg Ala Ser Val His  
 35 40 45  
 Gly Gln Leu Arg Ile Met Phe Pro Met Val Ala Leu Leu Lys Glu Phe  
 50 55 60  
 20 Arg Ala Ala Lys Ala Val Phe Asp Glu Glu Lys Ala Asn Leu Leu Ala  
 65 70 75 80  
 Glu Gly Val Ala Val Ala Asp Asn Ile Gln Val Gly Ile Met Ile Glu  
 85 90 95  
 25 Ile Pro Ala Ala Ala Met Leu Ala Asp Gln Phe Ala Lys Glu Val Asp  
 100 105 110  
 Phe Phe Ser Ile Gly Thr Asn Asp Leu Ile Gln Tyr Thr Met Ala Ala  
 115 120 125  
 30 Asp Arg Met Asn Glu Gln Val Ser Tyr Leu Tyr Gln Pro Tyr Asn Pro  
 130 135 140  
 35 Ser Ile Leu Arg Leu Ile Asn Asn Val Ile Lys Ala Ala His Ala Glu  
 145 150 155 160  
 Gly Lys Trp Ala Gly Met Cys Gly Glu Met Ala Gly Asp Gln Gln Ala  
 165 170 175  
 40 Val Pro Leu Leu Val Gly Met Gly Leu Asp Glu Phe Ser Met Ser Ala  
 180 185 190  
 Thr Cys Thr Ser Tyr Thr Gln Leu Asp Glu Glu Thr Arg His Ser  
 195 200 205

(2) INFORMATION FOR SEQ ID NO:178:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 203 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

60



-267-

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

5 Met Gln Met Ala Tyr Arg Cys Asn Leu Arg Asn Asn Gly Lys Arg Arg  
 1 5 10 15  
 Ile Gly Ile Arg Glu Met Thr Glu Met Leu Lys Gly Ile Ala Ala Ser  
 20 25 30  
 10 Asp Gly Val Ala Val Ala Lys Ala Tyr Leu Leu Val Gln Pro Asp Leu  
 35 40 45  
 Ser Phe Glu Thr Ile Thr Val Glu Asp Thr Asn Ala Glu Glu Ala Arg  
 50 55 60  
 15 Leu Asp Ala Ala Leu Gln Ala Ser Gln Asp Glu Leu Ser Val Ile Arg  
 65 70 75 80  
 20 Glu Lys Ala Val Gly Thr Leu Gly Glu Glu Ala Ala Gln Val Phe Asp  
 85 90 95  
 Ala His Leu Met Val Leu Ala Asp Pro Glu Met Ile Ser Gln Ile Lys  
 100 105 110  
 25 Glu Thr Ile Arg Ala Lys Lys Val Asn Ala Glu Ala Gly Leu Lys Glu  
 115 120 125  
 Val Thr Asp Met Phe Ile Thr Ile Phe Glu Gly Met Glu Asp Asn Pro  
 130 135 140  
 30 Tyr Met Gln Glu Arg Ala Arg Asp Ile Arg Asp Val Thr Lys Arg Val  
 145 150 155 160  
 35 Leu Ala Asn Leu Leu Gly Lys Lys Leu Pro Asn Pro Ala Ser Ile Asn  
 165 170 175  
 Glu Glu Val Ile Val Ile Ala His Asp Leu Thr Pro Ser Asp Thr Ala  
 180 185 190  
 40 Gln Leu Asp Lys Asn Phe Val Lys Ala Phe Val Thr Asn Ile Gly Gly  
 195 200 205  
 Arg Thr Ser His Ser Ala Ile Met Ala Arg Thr Leu Glu Ile Ala Ala  
 210 215 220  
 45 Val Leu Gly Thr Asn Asn Ile Thr Glu Ile Val Lys Asp Gly Asp Ile  
 225 230 235 240  
 50 Leu Ala Val Asn Gly Ile Thr Gly Glu Val Ile Ile Asn Pro Thr Asp  
 245 250 255  
 Glu Gln Ala Ala Glu Phe Lys Ala Ala Gly Glu Ala Tyr Ala Thr Lys  
 260 265 270  
 55 Ala Glu Trp Ala Leu Leu Lys Asp Ala Gln Gln  
 275 280

## (2) INFORMATION FOR SEQ ID NO:179:

60 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 168 amino acids

WO 98/26072

PCT/US97/22578

-268-

(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

5 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

15 Met Ile Gly Arg Leu Ala Pro Tyr Asp Lys Gly Gln Ile Ile Tyr Asp  
1 5 10 15  
Gly Thr Ser Leu Lys Asp Ile Lys Pro Ser Val Phe Phe Arg Asp Tyr  
20 25 30  
Leu Gly Tyr Leu Phe Gln Asp Phe Gly Leu Ile Glu Ser Gln Thr Val  
35 40 45  
Lys Glu Asn Leu Asn Leu Gly Leu Val Gly Lys Lys Leu Lys Glu Lys  
50 55 60  
25 Glu Lys Ile Ser Leu Met Lys Gln Ala Leu Asn Arg Val Asn Leu Ser  
65 70 75 80  
Tyr Leu Asp Leu Lys Gln Pro Ile Phe Glu Leu Ser Gly Gly Glu Ala  
30 85 90 95  
Gln Arg Val Ala Leu Ala Lys Ile Ile Leu Lys Asp Pro Pro Leu Ile  
100 105 110  
35 Leu Ala Asp Glu Pro Thr Ala Ser Leu Asp Pro Lys Asn Ser Glu Glu  
115 120 125  
Leu Leu Ser Ile Leu Glu Ser Leu Lys Asn Pro Asn Arg Thr Ile Ile  
40 130 135 140  
Ile Ala Thr His Asn Pro Leu Ile Trp Glu Gln Val Asp Gln Val Ile  
145 150 155 160  
45 Arg Val Thr Asp Leu Ser His Arg  
165

(2) INFORMATION FOR SEQ ID NO:180:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 210 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

55 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

60

-269-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

Met Lys Ala His Val Ser Tyr Leu Ser Met Gly Glu Lys Arg Phe Val  
1 5 10 15  
5 Tyr Asn Asn Gly Glu Asn Pro Val Ser Thr Gln Tyr Leu Thr Asp Pro  
20 25 30  
10 Ile Leu Val Val Phe Thr Pro Thr Ser Thr Gly Asp Ser Phe Ile Ser  
35 40 45  
Leu Ser Ser Trp Ser Ile Asn Ala Gly Lys Gln Leu Phe Ile Lys Gly  
50 55 60  
15 Tyr Glu Ser Gly Leu Glu Leu Leu Lys Lys Ala Gly Ile Tyr Glu Gln  
65 70 75 80  
Val Ser Tyr Leu Lys Glu Gly Arg Ser Val Tyr Leu Thr Arg Tyr Asn  
85 90 95  
20 Glu Val Gln Thr Glu Thr Ala Thr Leu Ile Leu Gly Ala Ile Val Gly  
100 105 110  
Ile Ala Ser Ser Leu Leu Leu Phe Tyr Ser Val Asn Leu Leu Tyr Phe  
115 120 125  
25 Glu Gln Phe Arg Arg Asp Ile Leu Ile Lys Arg Ile Ser Gly Leu Arg  
130 135 140  
30 Phe Phe Glu Thr His Ala Gln Tyr Met Val Ser Gln Phe Ala Ser Phe  
145 150 155 160  
Val Phe Gly Ala Ser Leu Phe Ile Leu Ser Ser Arg Asp Leu Val Ile  
165 170 175  
35 Gly Leu Leu Thr Leu Leu Val Phe Leu Ala Ser Ala Val Leu Thr Leu  
180 185 190  
Tyr Arg Gln Ala Gln Lys Glu Ser Arg Val Ser Met Thr Ile Met Lys  
195 200 205  
40 Gly Lys  
210

45 (2) INFORMATION FOR SEQ ID NO:181:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 227 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:



PCT/US97/22578

	Met	Ser	Lys	Asp	Lys	Lys	Asn	Glu	Asp	Lys	Glu	Thr	Leu	Glu	Glu	Leu
	1				5					10					15	
5	Lys	Glu	Leu	Ser	Glu	Trp	Gln	Lys	Arg	Asn	Gln	Glu	Tyr	Leu	Lys	Lys
				20					25					30		
	Lys	Ala	Glu	Glu	Glu	Val	Ala	Leu	Ala	Glu	Glu	Lys	Glu	Lys	Glu	Arg
			35					40					45			
10	Gln	Ala	Arg	Met	Gly	Glu	Glu	Ser	Glu	Lys	Ser	Glu	Asp	Lys	Gln	Asn
					50			55				60				
	Gln	Glu	Ser	Glu	Thr	Asp	Gln	Glu	Asp	Ser	Glu	Ser	Ala	Lys	Glu	Glu
					70						75				80	
15	Ser	Glu	Glu	Lys	Val	Ala	Ser	Ser	Glu	Ala	Asp	Lys	Glu	Lys	Glu	Glu
				85					90					95		
	Pro	Glu	Ser	Lys	Glu	Lys	Glu	Glu	Gln	Asp	Lys	Lys	Leu	Ala	Lys	Lys
				100					105					110		
	Ala	Thr	Lys	Glu	Lys	Pro	Ala	Lys	Ala	Lys	Ile	Pro	Gly	Ile	His	Ile
			115				120						125			
25	Leu	Arg	Ala	Phe	Thr	Ile	Leu	Phe	Pro	Ser	Leu	Leu	Leu	Ile	Val	
		130					135					140				
	Ser	Ala	Tyr	Leu	Leu	Ser	Pro	Tyr	Ala	Thr	Met	Lys	Asp	Ile	Arg	Val
		145				150					155				160	
30	Glu	Gly	Thr	Val	Gln	Thr	Thr	Ala	Asp	Asp	Ile	Arg	Gln	Ala	Ser	Gly
				165					170					175		
	Ile	Gln	Asp	Ser	Asp	Tyr	Thr	Ile	Asn	Leu	Leu	Leu	Asp	Lys	Ala	Lys
				180					185					190		
35	Tyr	Glu	Lys	Gln	Ile	Lys	Ser	Asn	Tyr	Trp	Val	Glu	Ser	Ala	Gln	Leu
			195					200					205			
40	Val	Tyr	Gln	Phe	Pro	Thr	Lys	Phe	Thr	Ile	Lys	Val	Lys	Glu	Tyr	Asp
		210					215					220				
	Ile	Val	Ala	Tyr	Tyr	Ile	Ser	Gly	Glu	Asn	His	Tyr	Pro	Ile	Leu	Ser
		225				230				235					240	
45	Ser	Gly	Gln	Leu	Glu	Thr	Ser	Ser	Val	Ser	Leu	Asn	Ser	Leu	Pro	Glu
				245					250					255		
50	Thr	Tyr	Leu	Ser	Val	Leu	Phe	Asn	Asp	Ser	Glu	Gln	Ile	Lys	Val	Phe
			260					265						270		
	Val	Ser	Glu	Leu	Ala	Gln	Ile	Ser	Pro	Glu	Leu	Lys	Ala	Ala	Ile	Gln
			275					280					285			
55	Lys	Val	Glu	Leu	Ala	Pro	Ser	Lys	Val	Thr	Ser	Asp	Leu	Ile	Arg	Leu
			290				295				300					
60	Thr	Met	Asn	Asp	Ser	Asp	Glu	Val	Leu	Val	Pro	Leu	Ser	Glu	Met	Ser
		305			310						315					320

WO 98/26072

PCT/US97/22578

-272-

Lys Lys Leu Pro Tyr Tyr Ser Lys Ile Lys Pro Gln Leu Ser Glu Pro  
 325 330 335  
 5 Ser Val Val Asp Met Glu Ala Gly Ile Tyr Ser Tyr Thr Val Ala Asp  
 340 345 350  
 Lys Leu Ile Met Glu Ala Glu Glu Lys Ala Lys Gln Glu Ala Lys Glu  
 355 360 365  
 10 Ala Glu Lys Lys Gln Glu Glu Glu Gln Lys Lys Gln Glu Glu Glu Ser  
 370 375 380  
 Asn Arg Asn Gln Thr Asn Gln Arg Ser Ser Arg Arg  
 385 390 395  
 15 (2) INFORMATION FOR SEQ ID NO:183:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 165 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant  
 20 (ii) MOLECULE TYPE: peptide  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:  
 Met Glu Arg Val Val Asp Ile Leu Lys Ala Glu Phe Asp Arg Ser Phe  
 1 5 10 15  
 35 Lys Leu Ile Asn Ser Lys Thr Tyr Pro Val Ser Gly Gly Glu Leu Asn  
 20 25 30  
 40 Pro Ala Asn Val Asp Ser Glu Ile Glu Ala Phe Ala Gln Leu Gly Val  
 35 40 45  
 Ser Arg Gly Leu Asp Ser Lys Glu Ala His Tyr Leu Ala Asn Leu Tyr  
 50 55 60  
 45 Gly Ser Asn Ala Pro Lys Val Phe Ala Leu Ala His Ser Leu Glu Gln  
 65 70 75 80  
 Ala Pro Gly Leu Ser Leu Ala Asp Thr Leu Ser Leu His Tyr Ala Met  
 85 90 95  
 50 Arg Asn Glu Leu Ala Leu Ser Pro Val Asp Phe Leu Leu Arg Arg Thr  
 100 105 110  
 55 Asn His Met Leu Phe Met Arg Asp Ser Leu Asp Ser Ile Val Glu Pro  
 115 120 125  
 Val Leu Asp Glu Met Gly Arg Phe Tyr Asp Trp Thr Glu Glu Glu Lys  
 130 135 140  
 60 Ala Thr Tyr Arg Ala Asp Val Glu Ala Ala Leu Ala Asn Asn Asp Leu  
 145 150 155 160

WO 98/26072

PCT/US97/22578

-273-

Ala Glu Leu Lys Asn  
165

5 (2) INFORMATION FOR SEQ ID NO:184:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 233 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

Met Asn Glu Leu Phe Gly Glu Phe Leu Gly Thr Leu Ile Leu Ile Leu  
1 5 10 15

25      Leu Gly Asn Gly Val Val Ala Gly Val Val Leu Pro Lys Thr Lys Ser  
                20                         25                         30

Asn Ser Ser Gly Trp Ile Val Ile Thr Met Gly Trp Gly Ile Ala Val  
35 40 45

Ala Val Ala Val Phe Val Ser Gly Lys Leu Ser Pro Ala His Leu Asn  
50 55 60

35 Pro Ala Val Thr Ile Gly Val Ala Leu Lys Gly Gly Leu Pro Trp Ala  
65 70 75 80

Ser Val Leu Pro Tyr Ile Leu Ala Gln Phe Ala Gly Ala Met Leu Gly  
85 90 95

40            Gln Ile Leu Val Trp Leu Gln Phe Lys Pro His Tyr Glu Ala Glu Glu  
                        100                         105                         110

Asn Ala Gly Asn Ile Leu Ala Thr Phe Ser Thr Gly Pro Ala Ile Lys  
115 120 125

Asp Thr Val Ser Asn Leu Ile Ser Glu Ile Leu Gly Thr Phe Val Leu  
130 135 140

50 Val Leu Thr Ile Phe Ala Leu Gly Leu Tyr Asp Phe Gln Ala Gly Ile  
145 150 155 160

Gly Thr Phe Ala Val Gly Thr Leu Ile Val Gly Ile Gly Leu Ser Leu  
165 170 175

55 Gly Gly Thr Thr Gly Tyr Ala Leu Asn Pro Ala Arg Asp Leu Gly Pro  
180 185 190

Arg Ile Met His Ser Ile Leu Pro Ile Pro Asn Lys Gly Asp Gly Asp  
195 200 205

60 Trp Ser Tyr Ala Trp Ile Pro Val Val Gly Pro Val Ile Gly Ala Ala

WO 98/26072

PCT/US97/22578

-274-

210 215 220

Leu Ala Val Leu Val Leu Ser Leu Phe  
225 230

5 (2) INFORMATION FOR SEQ ID NO:185:

(i) SEQUENCE CHARACTERISTICS:  
10 (A) LENGTH: 135 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

15 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

Thr Thr Asp Asn Val Ile Asp Leu Phe Glu His Ile Phe Lys Met Phe  
1 5 10 15  
Asn Glu Asn Ile Val Met Ala Gly Lys Val Asn Leu Leu Asn Phe Ala  
20 25 30  
Asn Leu Ala Ala Tyr Gln Phe Phe Asp Gln Pro Gln Lys Val Ala Leu  
35 40 45  
Glu Ile Arg Glu Gly Leu Arg Glu Asp Gln Met Gln Asn Val Arg Val  
50 55 60  
Ala Asp Gly Gln Glu Ser Cys Leu Ala Asp Leu Ala Val Ile Ser Ser  
65 70 75 80  
Lys Phe Leu Ile Pro Tyr Arg Gly Val Gly Ile Leu Ala Ile Ile Gly  
85 90 95  
Pro Val Asn Leu Asp Tyr Gln Gln Leu Ile Asn Gln Ile Asn Val Val  
100 105 110  
Asn Arg Val Leu Thr Met Lys Leu Thr Asp Phe Tyr Arg Tyr Leu Ser  
115 120 125  
Ser Asn His Tyr Glu Val His  
130 135

50 (2) INFORMATION FOR SEQ ID NO:186:

(i) SEQUENCE CHARACTERISTICS:  
55 (A) LENGTH: 153 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

60 (iii) HYPOTHETICAL: NO



WO 98/26072

PCT/US97/22578

-275-

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

5 Met Ile Ala Lys Glu Phe Glu Thr Phe Leu Leu Gly Gln Glu Glu Thr  
1 5 10  
10 Phe Leu Thr Pro Ala Lys Asn Leu Ala Val Leu Ile Asp Thr His Asn  
20 25 30  
Ala Asp His Ala Thr Leu Leu Ser Gln Met Thr Tyr Thr Arg Val  
35 40 45  
15 Pro Val Val Thr Asp Glu Lys Gln Phe Val Gly Thr Ile Gly Leu Arg  
50 55 60  
20 Asp Ile Met Ala Tyr Gln Met Glu His Asp Leu Ser Gln Glu Ile Met  
65 70 75 80  
Ala Asp Thr Asp Ile Val His Met Thr Lys Thr Asp Val Ala Val Val  
85 90 95  
25 Ser Pro Asp Phe Thr Ile Thr Glu Val Leu His Lys Leu Val Asp Glu  
100 105 110  
Ser Phe Leu Pro Val Val Asp Ala Glu Gly Ile Phe Gln Gly Ile Ile  
115 120 125  
30 Thr Arg Lys Ser Ile Leu Lys Ala Val Asn Ala Leu Leu His Asp Phe  
130 135 140  
Ser Lys Glu Tyr Glu Ile Arg Cys Gln  
145 150

(2) INFORMATION FOR SEQ ID NO:187:

(i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 173 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

45 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

55 Met Ala Lys Gln Thr Ile Ile Val Met Ser Asp Ser His Gly Asp Ser  
1 5 10  
Leu Ile Val Glu Glu Val Arg Asp Arg Tyr Val Gly Lys Val Asp Ala  
20 25 30  
60 Val Phe His Asn Gly Asp Ser Glu Leu Arg Pro Asp Ser Pro Leu Trp  
35 40 45



WO 98/26072

PCT/US97/22578

-277-

Arg Gly Val Asn Ser Ser Gly Ala Leu Ala Lys Leu Gln Ala Phe Asp  
115 120 125

5 Leu Ile Lys Glu Asp Gly Lys Lys Glu Val Leu Gly Arg Pro Asn Leu  
130 135 140

Tyr Val Thr Thr Asp Tyr Phe Leu Asp Tyr Met Gly Ile Asn His Leu  
145 150 155 160

10 Glu Glu Leu Pro Val Ile Asp Glu Leu Glu Ile Gln Ala Gln Glu Ser  
165 170 175

Gln Leu Phe Gly Glu Arg Ile Glu Glu Asp Glu Asn Gln  
180 185

15

(2) INFORMATION FOR SEQ ID NO:189:

(i) SEQUENCE CHARACTERISTICS:  
20 (A) LENGTH: 214 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

25 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

30 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:189:

35 Met Arg Asp Arg Ile Ser Ala Phe Leu Glu Lys Gln Gly Leu Ser  
1 5 10 15

Val Asn Ser Lys Gln Ser Tyr Lys Tyr Asp Leu Glu Gln Phe Leu Asp  
20 25 30

40 Met Val Gly Glu Arg Ile Ser Glu Thr Ser Leu Lys Ile Tyr Gln Ala  
35 40 45

Gln Leu Ala Asn Leu Lys Ile Ser Ala Gln Lys Arg Lys Ile Ser Ala  
50 55 60

45 Cys Asn Gln Phe Leu Tyr Phe Leu Tyr Gln Lys Gly Glu Val Asp Ser  
65 70 75 80

50 Phe Tyr Arg Leu Glu Leu Ala Lys Gln Ala Glu Lys Lys Thr Glu Lys  
85 90 95

Pro Glu Ile Leu Tyr Leu Asp Ser Phe Trp Gln Glu Ser Asp His Pro  
100 105 110

55 Glu Gly Arg Leu Leu Ala Leu Leu Ile Leu Glu Met Gly Leu Leu Pro  
115 120 125

Ser Glu Ile Leu Ala Ile Lys Val Ala Asp Ile Asn Leu Asp Phe Gln  
130 135 140

60 Val Leu Arg Ile Ser Lys Ala Ser Gln Gln Arg Ile Val Thr Ile Pro

PCT/US97/22578

```

145          150          155          160
Thr Ala Leu Leu Ser Glu Leu Glu Pro Leu Met Gly Gln Thr Tyr Leu
5          165
Phe Glu Arg Gly Glu Lys Pro Tyr Ser Arg Gln Trp Ala Phe Arg Gln
          180          185          190
Leu Glu Ser Phe Val Arg Arg Arg Phe Pro Ser Leu Ser Ala Gln Val
10          195          200          205
Leu Arg Asp Ser Leu Phe
          210

15 (2) INFORMATION FOR SEQ ID NO:190:

      (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 239 amino acids
            (B) TYPE: amino acid
20          (C) STRANDEDNESS: not relevant
            (D) TOPOLOGY: not relevant

      (ii) MOLECULE TYPE: peptide

25      (iii) HYPOTHETICAL: NO

      (iv) ANTI-SENSE: NO

30      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

Met Arg Ile Asn Lys Tyr Ile Ala His Ala Gly Val Ala Ser Arg Arg
1          5          10          15
Lys Ala Glu Glu Leu Ile Lys Gln Gly Leu Val Thr Val Asn Gly Gln
35          20          25          30
Val Val Arg Glu Leu Ala Thr Thr Ile Lys Ser Gly Asp Lys Val Glu
40          35          40          45
Val Glu Gly Gln Pro Ile Tyr Asn Glu Glu Lys Val Tyr Tyr Leu Leu
          50          55          60
Asn Lys Pro Arg Gly Val Ile Ser Ser Val Thr Asp Asp Lys Gly Arg
45          65          70          75          80
Lys Thr Val Val Asp Leu Leu Pro Asn Val Lys Glu Arg Ile Tyr Pro
          85          90          95
Val Gly Arg Leu Asp Trp Asp Thr Ser Gly Val Leu Ile Leu Thr Asn
50          100          105          110
Asp Gly Asp Phe Thr Asp Glu Met Ile His Pro Arg Asn Glu Ile Asp
55          115          120          125
Lys Val Tyr Val Ala Arg Val Lys Gly Val Ala Asn Lys Asp Asn Leu
          130          135          140
Arg Pro Leu Thr Arg Gly Leu Glu Ile Asp Gly Lys Lys Thr Lys Pro
60          145          150          155          160

```

WO 98/26072

PCT/US97/22578

-279-

Ala Val Tyr Glu Ile Leu Lys Val Asp Pro Val Lys Asn Arg Ser Val  
165 170 175

5 Val Gln Leu Thr Ile His Glu Gly Arg Asn His Gln Val Lys Lys Met  
180 185 190

Phe Glu Ala Val Gly Leu Gln Val Asp Lys Leu Ser Arg Thr Arg Phe  
195 200 205

10 Gly His Leu Asp Leu Thr Leu Arg Pro Gly Glu Ser Arg Arg Leu Asn  
210 215 220

Lys Lys Glu Ile Ser Gln Leu His Thr Met Ala Val Thr Lys Lys  
225 230 235

15 (2) INFORMATION FOR SEQ ID NO:191:

(i) SEQUENCE CHARACTERISTICS:  
20 (A) LENGTH: 243 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

25 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

Met Asp Ile Lys Leu Lys Arg Phe Leu Lys Asp Pro Gly Leu Ala Leu  
1 5 10 15

35 Cys Ile Trp Phe Leu Ser Thr Lys Met Asp Ile Tyr Asp Val Pro Ile  
20 25 30

40 Thr Glu Val Ile Glu Gln Tyr Leu Ala Tyr Val Ser Thr Leu Gln Ala  
35 40 45

Met Arg Leu Glu Val Thr Gly Glu Tyr Met Val Met Ala Ser Gln Leu  
50 55 60

45 Met Leu Ile Lys Ser Arg Lys Leu Leu Pro Lys Val Ala Glu Val Thr  
65 70 75 80

Asp Leu Gly Asp Asp Leu Glu Gln Asp Leu Leu Ser Gln Ile Glu Glu  
85 90 95

50 Tyr Arg Lys Phe Lys Leu Leu Gly Glu His Leu Glu Ala Lys His Gln  
100 105 110

55 Glu Arg Ala Gln Tyr Tyr Ser Lys Ala Pro Thr Glu Leu Ile Tyr Glu  
115 120 125

Asp Ala Glu Leu Val His Asp Lys Thr Thr Ile Asp Leu Phe Leu Ala  
130 135 140

60 Phe Ser Asn Ile Leu Ala Lys Lys Lys Glu Glu Phe Ala Gln Asn His  
145 150 155 160

WO 98/26072

PCT/US97/22578

-280-

Thr Thr Ile Leu Arg Asp Glu Tyr Lys Ile Glu Asp Met Met Ile Ile  
 165 170 175  
 5 Val Lys Glu Ser Leu Ile Gly Arg Asp Gln Leu Arg Leu Gln Asp Leu  
 180 185 190  
 Phe Lys Glu Ala Gln Asn Val Gln Glu Val Ile Thr Leu Phe Leu Ala  
 195 200 205  
 10 Thr Leu Glu Leu Ile Lys Thr Gln Glu Leu Ile Leu Val Gln Glu Glu  
 210 215 220  
 Ser Phe Gly Asp Ile Tyr Leu Met Glu Lys Lys Glu Glu Ser Gln Val  
 225 230 235 240  
 Pro Gln Ser  
 20 (2) INFORMATION FOR SEQ ID NO:192:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 336 amino acids  
 (B) TYPE: amino acid  
 25 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant  
 (ii) MOLECULE TYPE: peptide  
 30 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:192:  
 Met Ala Gly Lys Arg Asp Ser Cys Gly Ala Cys Arg Ile Met Thr Asn  
 1 5 10 15  
 40 Lys Ile Tyr Glu Tyr Lys Asp Asp Gln Asn Trp Tyr Val Gly Ser Tyr  
 20 25 30  
 Ser Ile Phe Gly Gly Val Asn Ser Leu Ser Asp Tyr Lys Ala Asp Phe  
 35 40  
 45 Pro Leu Phe Glu Phe Ser Lys Ile Phe Gly Asp Glu Tyr Gly Phe  
 50 55 60  
 50 Pro Leu Ser Val Thr Val Leu Arg Tyr Gly Ser Thr Tyr Arg Leu Phe  
 65 70 75 80  
 Ser Phe Val Val Asp Met Leu Asn Gln Glu Met Gly Arg Asn Leu Glu  
 85 90 95  
 55 Val Ile Gln Arg His Gly Ala Leu Leu Val Glu Asn Gly Gln Leu  
 100 105 110  
 Leu Tyr Val Glu Leu Pro Lys Glu Gly Val Asn Val His Asp Phe Phe  
 115 120 125  
 60 Glu Thr Ser Lys Val Arg Glu Thr Leu Leu Ile Ala Thr Arg Asn Glu

WQ 98/26072

PCT/US97/22578

-281-

	130	135	140
5	Gly Lys Thr Lys Glu Phe Arg Ala Ile Phe Asp Lys Leu Gly Tyr Asp 145 150 155 160		
	Val Glu Asn Leu Asn Asp Tyr Pro Asp Leu Pro Glu Val Ala Glu Thr 165 170 175		
10	Gly Met Thr Phe Glu Glu Asn Ala Arg Leu Lys Ala Glu Thr Ile Ser 180 185 190		
	Gln Leu Thr Gly Lys Met Val Leu Ala Asp Asp Ser Gly Leu Lys Val 195 200 205		
15	Asp Val Leu Gly Gly Leu Pro Gly Val Trp Ser Ala Arg Phe Ala Gly 210 215 220		
	Val Gly Ala Thr Asp Arg Glu Asn Asn Ala Lys Leu Leu His Glu Leu 225 230 235 240		
20	Ala Met Val Phe Glu Leu Lys Asp Arg Ser Ala Gln Phe His Thr Thr 245 250 255		
	Leu Val Val Ala Ser Pro Asn Lys Glu Ser Leu Val Val Glu Ala Asp 260 265 270		
25	Trp Ser Gly Tyr Ile Asn Phe Glu Pro Lys Gly Glu Asn Gly Phe Gly 275 280 285		
30	Tyr Asp Pro Leu Phe Leu Val Gly Glu Thr Gly Glu Ser Ser Ala Glu 290 295 300		
	Leu Thr Leu Glu Glu Lys Asn Ser Gln Ser His Arg Ala Leu Ala Val 305 310 315 320		
35	Lys Lys Leu Leu Glu Val Phe Pro Ser Trp Gln Ser Lys Pro Ser Leu 325 330 335		

40 (2) INFORMATION FOR SEQ ID NO:193:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 219 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:193:

60	Glu Asn Asn Tyr Glu Pro Gln Tyr Ile Asn Ile Arg Gly Lys Gly Pro 1 5 10 15
	Leu Ile Asn Asp Leu Lys Lys Glu Ala Lys Lys Ala Asn Lys Val Phe 20 25 30

WO 98/26072

PCT/US97/22578

-282-

Leu Ala Ser Asp Pro Asp Arg Glu Gly Glu Ala Ile Ser Trp His Leu  
35 40 45

5 Ala His Ile Leu Asn Leu Asp Glu Asn Asp Ala Asn Arg Val Val Phe  
50 55 60

Asn Glu Ile Thr Lys Asp Ala Val Lys Asn Ala Phe Lys Glu Pro Arg  
65 70 75 80

10 Lys Ile Asp Met Asp Leu Val Asp Ala Gln Gln Ala Arg Arg Ile Leu  
85 90 95

Asp Arg Leu Val Gly Tyr Ser Ile Ser Pro Ile Leu Trp Lys Val  
100 105 110

15 Lys Lys Gly Leu Ser Ala Gly Arg Val Gln Ser Ile Ala Leu Lys Leu  
115 120 125

20 Ile Ile Asp Arg Glu Asn Glu Ile Asn Ala Phe Gln Pro Glu Glu Tyr  
130 135 140

Trp Thr Val Asp Ala Val Phe Lys Lys Gly Thr Lys Gln Phe His Ala  
145 150 155 160

25 Ser Phe Tyr Gly Val Asp Gly Lys Lys Met Lys Leu Thr Ser Asn Asn  
165 170 175

Glu Val Lys Glu Val Leu Ser Arg Leu Thr Ser Lys Asp Phe Ser Val  
180 185 190

30 Asp Gln Val Asp Lys Lys Glu Arg Lys Ala Asn Ala Pro Leu Pro Tyr  
195 200 205

35 Thr Thr Ser Ser Met Gln Met Gly Cys Cys Gln  
210 215

(2) INFORMATION FOR SEQ ID NO:194:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 236 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:

Met Ser Ile His Ile Ala Ala Gln Gln Gly Glu Ile Ala Asp Lys Ile  
1 5 10 15

Leu Leu Pro Gly Asp Pro Leu Arg Ala Lys Phe Ile Ala Glu Asn Phe  
20 25 30

60 Leu Gly Asp Ala Val Cys Phe Asn Glu Val Arg Asn Met Phe Gly Tyr



WO 98/26072

PCT/US97/22578

-263-

	35	40	45
5	Thr Gly Thr Tyr Lys Gly His Arg Val Ser Val Met Gly Thr Gly Met 50 55 60		
	Gly Met Pro Ser Ile Ser Ile Tyr Ala Arg Glu Leu Ile Val Asp Tyr 65 70 80		
10	Gly Val Lys Lys Leu Ile Arg Val Gly Thr Ala Gly Ser Leu Asn Glu 85 90 95		
	Glu Val His Val Arg Glu Leu Val Leu Ala Gln Ala Ala Thr Asn 100 105 110		
15	Ser Asn Ile Val Arg Asn Asp Trp Pro Gln Tyr Asp Phe Pro Gln Ile 115 120 125		
	Ala Ser Phe Asp Leu Leu Asp Lys Ala Tyr His Ile Ala Lys Glu Leu 130 135 140		
20	Gly Met Thr Thr His Val Gly Asn Val Leu Ser Ser Asp Val Phe Tyr 145 150 155 160		
25	Ser Asn Tyr Phe Glu Lys Asn Ile Glu Leu Gly Lys Trp Gly Val Lys 165 170 175		
	Ala Val Glu Met Glu Ala Ala Ala Leu Tyr Tyr Leu Ala Ala Gln Tyr 180 185 190		
30	His Val Asp Ala Leu Ala Ile Met Thr Ile Ser Asp Ser Leu Val Asn 195 200 205		
	Pro Asp Glu Asp Thr Thr Ala Glu Glu Arg Gln Asn Thr Phe Thr Asp 210 215 220		
35	Met Met Lys Val Gly Leu Glu Thr Leu Ile Ala Glu 225 230 235		

## (2) INFORMATION FOR SEQ ID NO:195:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 477 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:195:

55	Ile Ile Phe Pro Ile Leu Thr Gly Thr Tyr Val Ala Arg Val Leu Asp 1 5 10 15
60	Arg Thr Asp Tyr Gly Tyr Phe Asn Ser Val Asp Thr Ile Leu Ser Phe 20 25 30

WO 98/26072

PCT/US97/22578

-284-

Phe Leu Pro Phe Ala Thr Tyr Gly Val Tyr Asn Tyr Gly Leu Arg Ala  
35 40 45

5 Ile Ser Asn Val Lys Asp Asn Lys Lys Asp Leu Asn Arg Thr Phe Ser  
50 55 60

Ser Leu Phe Tyr Leu Cys Ile Ala Cys Thr Ile Leu Thr Thr Ala Val  
65 70 75 80

10 Tyr Ile Leu Ala Tyr Pro Leu Phe Phe Thr Asp Asn Pro Ile Val Lys  
85 90 95

15 Lys Val Tyr Leu Val Met Gly Ile Gln Leu Ile Ala Gln Ile Phe Ser  
100 105 110

Ile Glu Trp Val Asn Glu Ala Leu Glu Asn Tyr Ser Phe Leu Phe Tyr  
115 120 125

20 Lys Thr Ala Phe Ile Arg Ile Leu Met Leu Val Ser Ile Phe Leu Phe  
130 135 140

Val Lys Asn Glu His Asp Ile Val Val Tyr Thr Leu Val Met Ser Leu  
145 150 155 160

25 Ser Thr Leu Ile Asn Tyr Leu Ile Ser Tyr Phe Trp Ile Lys Arg Asp  
165 170 175

Ile Lys Leu Val Lys Ile His Leu Ser Asp Phe Lys Pro Leu Phe Leu  
180 185 190

30 Pro Leu Thr Ala Met Leu Val Phe Ala Asn Ala Asn Met Leu Phe Thr  
195 200 205

35 Phe Leu Asp Arg Leu Phe Leu Val Lys Thr Gly Ile Asp Val Asn Val  
210 215 220

Ser Tyr Tyr Thr Ile Ala Gln Arg Ile Val Thr Val Ile Ala Gly Val  
225 230 235 240

40 Val Thr Gly Ala Ile Gly Val Ser Val Pro Arg Leu Ser Tyr Tyr Leu  
245 250 255

Gly Lys Gly Asp Lys Glu Ala Tyr Val Ser Leu Val Asn Arg Gly Ser  
260 265 270

45 Arg Ile Phe Asn Phe Phe Ile Ile Pro Leu Ser Phe Gly Leu Met Val  
275 280 285

50 Leu Gly Pro Asn Ala Ile Leu Leu Tyr Gly Ser Glu Lys Tyr Ile Gly  
290 295 300

Gly Gly Ile Leu Thr Ser Leu Phe Ala Phe Arg Thr Ile Ile Leu Ala  
305 310 315 320

55 Leu Asp Thr Ile Leu Gly Ser Gln Ile Leu Phe Thr Asn Gly Tyr Glu  
325 330 335

Lys Arg Ile Thr Val Tyr Thr Val Phe Ala Gly Leu Leu Asn Leu Gly  
340 345 350

60 Leu Asn Ser Leu Leu Phe Phe Asn His Ile Val Ala Pro Glu Tyr Tyr

WO 98/26072

PCT/US97/22578

-285-

355 360 365

Leu Leu Thr Thr Met Leu Ser Glu Thr Ser Leu Leu Val Phe Tyr Ile  
370 375 380

5 Ile Phe Ile His Arg Lys Gln Leu Ile His Leu Gly His Ile Phe Ser  
385 390 395 400

Tyr Thr Val Arg Tyr Ser Leu Phe Ser Leu Ser Phe Val Ala Ile Tyr  
405 410 415

10 Phe Leu Ile Asn Phe Val Tyr Pro Val Asp Met Val Ile Asn Leu Pro  
420 425 430

15 Phe Leu Ile Asn Thr Gly Leu Ile Val Leu Leu Ser Ala Ile Ser Tyr  
435 440 445

Ile Ser Leu Leu Val Phe Thr Lys Asp Ser Ile Phe Tyr Glu Phe Leu  
450 455 460

20 Asn His Val Leu Ala Leu Lys Asn Lys Phe Lys Lys Ser  
465 470 475

(2) INFORMATION FOR SEQ ID NO:196:

25 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 148 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
30 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

35 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:

40 Phe Pro Ile Asp Arg Phe Asp Asp Pro Lys Val Ile Asp Thr Cys Tyr  
1 5 10 15

Lys Leu Glu Ser Phe Lys Leu Leu Ser Phe Ser Lys His Lys Asn Ile  
20 25 30

Val Tyr Lys Asp Ser Leu Leu Lys Asp Trp Ile Arg Thr Ala Phe Trp  
35 40 45

50 Leu Leu Leu Arg Pro Val Ser Pro Arg Tyr Phe Ala Asn Lys Ile Glu  
50 55 60

Lys Glu Ile Gln Lys Tyr Ser Arg Glu Asn Gly Gln Tyr Met Ala Phe  
65 70 75 80

55 Ile Pro Ser Lys Phe Lys Glu Lys Glu Val Phe Pro Ser Gly Thr Phe  
85 90 95

60 Asp Lys Thr Ile Asp Leu Pro Phe Glu Asn Leu Ser Leu Pro Ala Pro  
100 105 110

WO 98/26072

PCT/US97/22578

-286-

Glu Lys Phe Asp Thr Ile Leu Thr Gln Phe Tyr Gly Asp Tyr Met Thr  
 115 120  
 5 Leu Pro Pro Glu Glu Lys Arg Phe Tyr Ser His Glu Phe His Ala Tyr  
 130 135 140  
 Lys Leu Glu Asp  
 145

10 (2) INFORMATION FOR SEQ ID NO:197:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 280 amino acids  
 (B) TYPE: amino acid  
 15 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant  
 (ii) MOLECULE TYPE: peptide  
 20 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:  
 Met Asn Phe Thr Leu Ile Asn Trp Arg Ile Arg Met Gln Tyr Leu Glu  
 1 5 10 15  
 30 Lys Lys Glu Ile Lys Glu Ile Gln Leu Ala Leu Leu Asp Tyr Ile Asp  
 20 25 30  
 Glu Thr Cys Lys Lys His Asp Ile Pro Tyr Phe Leu Ser Tyr Gly Thr  
 35 35 40 45  
 Met Leu Gly Ala Ile Arg His Lys Gly Met Ile Pro Trp Asp Asp Asp  
 50 55 60  
 40 Ile Asp Ile Ser Leu Tyr Arg Glu Asp Tyr Glu Arg Leu Leu Lys Ile  
 65 70 75 80  
 Ile Glu Glu Glu Asn His Pro Arg Tyr Lys Val Leu Ser Tyr Asp Thr  
 85 90 95  
 45 Ser Ser Trp Tyr Phe His Asn Phe Ala Ser Ile Leu Asp Thr Ser Thr  
 100 105 110  
 Val Ile Glu Asp His Val Lys Tyr Lys Arg His Asp Thr Ser Leu Phe  
 115 120 125  
 50 Ile Asp Val Phe Pro Ile Asp Arg Phe Thr Asp Leu Ser Ile Val Asp  
 130 135 140  
 55 Lys Ser Tyr Lys Tyr Val Ala Leu Arg Gln Leu Ala Tyr Ile Lys Lys  
 145 150 155 160  
 Ser Arg Ala Val His Gly Asp Ser Lys Leu Lys Asp Phe Leu Arg Leu  
 165 170 175  
 60 Cys Ser Trp Tyr Ala Leu Arg Phe Val Asn Pro Arg Tyr Phe Tyr Lys  
 180 185 190

WO 98/26072

PCT/US97/22578

-267-

Lys Ile Asp Gln Leu Val Lys Asn Ala Val Thr Asn Thr Pro Gln Tyr  
 195 200 205  
 5 Glu Gly Gly Val Gly Ile Gly Lys Glu Gly Met Lys Glu Ile Phe Pro  
 210 215 220  
 Val Asp Thr Phe Lys Glu Leu Ile Leu Thr Glu Phe Glu Gly Arg Met  
 225 230 235 240  
 10 Leu Pro Val Pro Lys Lys Tyr Asp Gln Phe Leu Thr Gln Met Tyr Gly  
 245 250 255  
 Asp Tyr Met Thr Pro Pro Ser Lys Glu Met Gln Glu Trp Tyr Ser His  
 260 265 270  
 15 Ser Ile Lys Ala Tyr Arg Lys Asn  
 275 280  
 20 (2) INFORMATION FOR SEQ ID NO:198:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 223 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant  
 25  
 (ii) MOLECULE TYPE: peptide  
 30 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:  
 Lys Gly Phe Ile Pro Trp Asp Asp Asp Leu Asp Phe Phe Met Pro Arg  
 1 5 10 15  
 40 Lys Asp Tyr Glu Lys Leu Ala Glu Leu Trp Pro Arg Tyr Ala Asp Glu  
 20 25 30  
 Arg Tyr Phe Leu Ser Lys Ser His Lys Asp Phe Val Asp Arg Asn Leu  
 35 40 45  
 45 Phe Ile Thr Ile Arg Asp Lys Lys Thr Thr Cys Ile Lys Pro Tyr Gln  
 50 55 60  
 50 Gln Asp Leu Asp Leu Pro His Gly Leu Ala Leu Asp Val Leu Pro Leu  
 65 70 75 80  
 Asp Tyr Tyr Pro Lys Asn Pro Ala Glu Arg Lys Lys Gln Val Arg Trp  
 85 90 95  
 55 Ala Leu Ile Tyr Ser Leu Phe Cys Ala Gln Thr Ile Pro Glu Lys His  
 100 105 110  
 Gly Asp Leu Met Lys Trp Gly Ser Arg Ile Leu Leu Gly Leu Thr Pro  
 115 120 125  
 60 Lys Ser Leu Arg Tyr Arg Ile Trp Lys Lys Ala Glu Lys Glu Met Thr

WO 98/26072

PCT/US97/22578

-286-

	130		135		140
	Lys Tyr Asp Leu Ala Asp	Cys Asp Gly Ile Thr	Glu Leu Cys Ser Gly		
	145	150	155	160	
5	Pro Gly Tyr Met Arg Asn	Lys Tyr Pro Ile Thr	Ser Phe Glu Asp Asn		
	165	170	175		
	Leu Phe Leu Pro Phe Glu	Gly Thr Glu Met Pro Ile	Pro Ile Gly Tyr		
10	180	185	190		
	Asp Val Tyr Leu Arg Thr	Ala Phe Gly Asp Tyr Met	Thr Pro Pro Pro		
	195	200	205		
15	Ala Asp Lys Gln Val Pro	His His Asp Thr Val Thr	Ala Asp Met		
	210	215	220		

(2) INFORMATION FOR SEQ ID NO:199:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 835 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:199:

35	Gly Phe Asp Asp Tyr His	Pro Ser Cys Gly Arg Ile Leu Ser Val Val
	1	5 10 15
	Thr Ser Gly Gly Glu Asp Ile Ala Asp	Ala Ile Ile Ile Leu Ala Val
40	20	25 30
	Val Ile Ile Asn Ala Ala Phe Gly Val Tyr Gln Glu Gly Lys Ala Glu	
	35	40 45
45	Glu Ala Ile Glu Ala Leu Lys Ser Met Ser Ser Pro Val Ala Arg Val	
	50	55 60
	Leu Arg Asp Gly His Met Ala Glu Ile Asp Ser Lys Glu Leu Val Pro	
	65	70 75 80
50	Gly Asp Ile Val Ala Leu Glu Ala Gly Asp Val Val Pro Ala Asp Leu	
	85	90 95
	Arg Leu Ile Glu Ala Asn Ser Leu Lys Ile Glu Glu Ala Ala Leu Thr	
55	100	105 110
	Gly Glu Ser Val Pro Val Glu Lys Asp Leu Ser Val Asp Leu Ala Thr	
	115	120 125
60	Asp Ala Gly Ile Gly Asp Arg Val Asn Met Ala Phe Gln Asn Ser Asn	
	130	135 140

WO 98/26072

PCT/US97/22578

-289-

	Val Thr Tyr Gly Arg Gly Met Gly Val Val Val Asn Thr Gly Met Tyr	145	150	155	160
5	Thr Glu Val Gly His Ile Ala Gly Met Leu Gln Asp Ala Asp Glu Thr	165	170	175	
	Asp Thr Pro Leu Lys Gln Asn Leu Asn Asn Leu Ser Lys Val Leu Thr	180	185	190	
10	Tyr Ala Ile Leu Val Ile Ala Leu Val Thr Phe Val Val Gly Val Phe	195	200	205	
	Ile Gln Gly Lys Asn Pro Leu Gly Glu Leu Leu Thr Ser Val Ala Leu	210	215	220	
15	Ala Val Ala Ala Ile Pro Gly Glu Leu Pro Ala Ile Val Thr Ile Val	225	230	235	240
	Leu Ser Leu Gly Thr Gln Val Leu Ala Lys Arg His Ser Ile Val Arg	245	250	255	
20	Lys Leu Pro Ala Val Glu Thr Leu Gly Ser Thr Glu Ile Ile Ala Ser	260	265	270	
	Asp Lys Thr Gly Thr Leu Thr Met Asn Lys Met Thr Val Glu Lys Val	275	280	285	
	Phe Tyr Asp Ala Val Leu His Asp Ser Ala Asp Asp Ile Glu Leu Gly	290	295	300	
30	Leu Glu Met Pro Leu Leu Arg Ser Val Val Leu Ala Asn Asp Thr Lys	305	310	315	320
	Ile Asp Val Glu Gly Asn Leu Ile Gly Asp Pro Thr Glu Thr Ala Phe	325	330	335	
35	Ile Gln Tyr Ala Leu Asp Lys Gly Tyr Asp Val Lys Gly Phe Leu Glu	340	345	350	
	Lys Tyr Pro Arg Val Ala Glu Leu Pro Phe Asp Ser Asp Arg Lys Leu	355	360	365	
	Met Ser Thr Val His Pro Leu Pro Asp Ser Arg Phe Leu Val Ala Val	370	375	380	
45	Lys Gly Ala Pro Asp Gln Leu Leu Lys Arg Cys Leu Leu Arg Asp Lys	385	390	395	400
	Ala Gly Asp Ile Ala Pro Ile Asp Glu Lys Val Thr Asn Leu Ile His	405	410	415	
	Thr Asn Asn Ser Glu Met Ala His Gln Ala Leu Arg Val Leu Ala Gly	420	425	430	
55	Ala Tyr Lys Ile Ile Asp Ser Ile Pro Glu Asn Leu Thr Ser Glu Glu	435	440	445	
	Leu Glu Asn Asp Leu Ile Phe Thr Gly Leu Ile Gly Met Ile Asp Pro	450	455	460	
60	Glu Arg Pro Glu Ala Ala Glu Ala Val Arg Val Ala Lys Glu Ala Gly				

WO 9826072

PCT/US97/22578

-290-

	465		470		475		480
	Ile Arg Pro Ile Met Ile Thr Gly Asp His Gln Asp Thr Ala Glu Ala						
		485			490		495
5	Ile Ala Lys Arg Leu Gly Ile Ile Asp Ala Asn Asp Thr Glu Gly His						
		500			505		510
10	Val Leu Thr Thr Gly Ala Glu Leu Asn Glu Leu Ser Asp Glu Glu Phe Glu						
		515			520		525
	Lys Val Val Gly Gln Tyr Ser Val Tyr Ala Arg Val Ser Pro Glu His						
		530			535		540
15	Lys Val Arg Ile Val Lys Ala Trp Gln Lys Gln Gly Lys Val Val Ala						
		545			550		555
	Met Thr Gly Asp Gly Val Asn Asp Ala Pro Ala Leu Lys Thr Ala Asp						
		565			570		575
20	Ile Gly Ile Gly Met Gly Ile Thr Gly Thr Glu Val Ser Lys Gly Ala						
		580			585		590
	Ser Asp Met Ile Leu Ala Asp Asp Asn Phe Ala Thr Ile Ile Val Ala						
		595			600		605
25	Val Glu Glu Gly Arg Lys Val Phe Ser Asn Ile Gln Lys Thr Ile Gln						
		610			615		620
30	Tyr Leu Leu Ser Ala Asn Thr Ala Glu Val Leu Thr Ile Phe Leu Ser						
		625			630		635
	Thr Leu Phe Gly Trp Asp Val Leu Gln Pro Val His Leu Leu Trp Ile						
		645			650		655
35	Asn Leu Val Thr Asp Thr Phe Pro Ala Ile Ala Leu Gly Val Glu Pro						
		660			665		670
40	Ala Glu Pro Gly Val Met Asn His Lys Pro Arg Gly Arg Lys Ala Ser						
		675			680		685
	Phe Phe Ser Gly Gly Val Leu Ser Ser Ile Ile Tyr Gln Gly Val Leu						
		690			695		700
45	Gln Ala Ala Leu Val Met Ser Val Tyr Gly Leu Ala Ile Ala Tyr Pro						
		705			710		715
	Val His Val Gly Asp Asn His Ala Ile His Ala Asp Ala Leu Thr Met						
		725			730		735
50	Ala Phe Ala Thr Leu Gly Leu Ile Gln Leu Phe His Ala Tyr Asn Val						
		740			745		750
55	Lys Ser Val Tyr Gln Ser Ile Leu Thr Val Gly Pro Phe Lys Ser Lys						
		755			760		765
	Thr Phe Asn Trp Ser Ile Leu Val Ser Phe Ile Leu Leu Met Ala Thr						
		770			775		780
60	Ile Val Val Glu Pro Leu Glu Gly Ile Phe His Val Thr Lys Leu Asp						
		785			790		795
							800



WO 98/26072

PCT/US97/22578

-291-

Leu Ser Gln Trp Gly Ile Val Met Ala Gly Ser Phe Ser Met Ile Ile  
805 810 815

5 Ile Val Glu Ile Val Lys Phe Ile Gln Arg Lys Leu Gly Phe Asp Lys  
820 825 830

Asn Ala Ile  
835

10 (2) INFORMATION FOR SEQ ID NO:200:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 525 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

15 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:200:

Gly Phe Ile Leu Phe Phe Val Leu Leu Gly Ala Val Phe Glu Glu Lys  
1 5 10 15

30 Met Arg Lys Asn Thr Ser Gln Ala Val Glu Lys Leu Leu Asp Leu Gln  
20 25 30

35 Ala Lys Thr Ala Glu Val Leu Ser Asp Asp Ser Tyr Val Gln Val Pro  
35 40 45

Leu Glu Gln Val Lys Val Gly Asp Leu Ile Arg Val Arg Pro Gly Glu  
50 55 60

40 Lys Ile Ala Val Asp Gly Val Val Val Glu Gly Val Ser Ser Ile Asp  
65 70 75 80

Glu Ser Met Val Thr Gly Glu Ser Leu Pro Val Asp Lys Thr Val Gly  
85 90 95

45 Asp Thr Val Ile Gly Ser Thr Ile Asn His Ser Gly Thr Leu Val Phe  
100 105 110

50 Arg Ala Glu Lys Val Gly Ser Glu Thr Val Leu Ala Gln Ile Val Asp  
115 120 125

Phe Val Lys Lys Ala Gln Thr Ser Arg Ala Pro Ile Gln Asp Leu Thr  
130 135 140

55 Asp Lys Ile Ser Gly Ile Phe Val Pro Val Val Ile Leu Gly Ile  
145 150 155 160

Met Thr Phe Trp Val Trp Phe Val Leu Leu Arg Asp Ser Val Val Val  
165 170 175

60 Leu Gly Ala Ser Phe Val Ser Ser Leu Leu Tyr Gly Val Ala Val Leu

WO 98/26072

PCT/US97/22578

-292-

	180	185	190
5	Ile Ile Ala Cys Pro Cys Ala 195	Leu Gly Leu Ala Thr 200	Pro Thr Ala Leu 205
	Met Val Gly Thr Gly Arg 210	Ser Ala Lys Met Gly 215	Val Leu Leu Lys Asn 220
10	Gly Thr Val Leu Gln Glu 225	Ile Gln Lys Val Gln 230	Thr Leu Val Phe Asp 235
	Lys Thr Gly Thr Leu Thr 245	Glu Gly Lys Pro Val 250	Val Thr Asp Ile Ile 255
15	Gly Asp Glu Val Glu Val 260	Phe Gly Leu Ala Ala 265	Ser Leu Glu Asp Ala 270
	Ser Gln His Pro Leu Ala 275	Glu Ala Ile Val Lys 280	Arg Ala Ser Glu Ala 285
20	Gly Leu Glu Phe Gln Thr 290	Val Glu Asn Phe Gln 295	Ala Leu His Gly Lys 300
	Gly Val Ser Gly Arg Ile 305	Asn Gly Lys Gln Val 310	Leu Leu Gly Asn Ala 315
25	Lys Met Leu Asp Gly Met 325	Asp Ile Ser Asn Thr 330	Tyr Gln Asp Lys Leu 335
30	Glu Glu Leu Glu Lys Glu 340	Ala Lys Thr Val Val 345	Phe Leu Ala Val Asp 350
	Asn Glu Ile Lys Lys Glu 355	Leu Leu Ala Leu Gln 360	Asp Ile Pro Lys Glu Asn 365
35	Ala Lys Leu Ala Ile Ser 370	Gln Leu Lys Lys Arg 375	Gly Leu Arg Thr Val 380
40	Met Leu Thr Gly Asp Asn 385	Ala Gly Val Ala Arg 390	Ala Ile Ala Asp Gln 395
	Ile Gly Ile Glu Glu Val 405	Ile Ala Gly Val Leu 410	Pro Glu Glu Lys Ala 415
45	His Glu Ile His Lys Leu 420	Gln Ala Ala Gly Lys 425	Val Ala Phe Val Gly 430
	Asp Gly Ile Asn Asp Ala 435	Pro Ala Leu Ser Val 440	Ala Asp Val Gly Ile 445
50	Ala Met Gly Ala Gly Thr 450	Asp Ile Ala Ile Glu 455	Ser Ala Asp Leu Val 460
55	Leu Thr Thr Asn Asn Leu 465	Leu Gly Val Val Arg 470	Ala Phe Asp Met Ser 475
	Lys Lys Thr Phe His Arg 485	Ile Leu Leu Asn Leu 490	Phe Trp Ala Phe Ile 495
60	Tyr Asn Val Val Gly Ile 500	Pro Ile Ala Ala Gly 505	Val Phe Ser Gly Val 510

WO 98/26072

PCT/US97/22578

-293-

Gly Trp Leu Ser Thr Gln Ile Gly Lys Ala Ser Pro Met  
515 520 525

5 (2) INFORMATION FOR SEQ ID NO:201:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 362 amino acids  
(B) TYPE: amino acid  
10 (C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

15 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:201:

Asn Asp Ile Ile Glu Phe Met Asp Lys Asn Lys Ile Met Gly Leu Thr  
1 5 10 15

25 Gln Arg Glu Val Lys Glu Arg Gln Ala Glu Gly Leu Val Asn Asp Phe  
20 25 30

Thr Ala Ser Ala Ser Thr Ser Thr Trp Gln Ile Val Lys Arg Asn Val  
35 40 45

30 Phe Thr Leu Phe Asn Ala Leu Asn Phe Ala Ile Ala Leu Ala Leu Ala  
50 55 60

35 Phe Val Gln Ala Trp Ser Asn Leu Val Phe Phe Ala Val Ile Cys Phe  
65 70 75 80

Asn Ala Phe Ser Gly Ile Val Thr Glu Leu Arg Ala Lys His Met Val  
85 90 95

40 Asp Lys Leu Asn Leu Met Thr Lys Glu Lys Val Lys Thr Ile Arg Asp  
100 105 110

Gly Gln Glu Val Ala Leu Asn Pro Glu Glu Leu Val Leu Gly Asp Val  
115 120 125

45 Ile Arg Leu Ser Ala Gly Glu Gln Ile Pro Ser Asp Ala Leu Val Leu  
130 135 140

50 Glu Gly Phe Ala Glu Val Asn Glu Ala Met Leu Thr Gly Glu Ser Asp  
145 150 155 160

Leu Val Gln Lys Glu Val Asp Gly Leu Leu Leu Ser Gly Ser Phe Leu  
165 170 175

55 Ala Ser Gly Ser Val Leu Ser Gln Val His Val Gly Ala Asp Asn  
180 185 190

Tyr Ala Ala Lys Leu Met Leu Glu Ala Lys Thr Val Lys Pro Ile Asn  
195 200 205

60 Ser Arg Ile Met Lys Ser Leu Asp Lys Leu Ala Gly Phe Thr Gly Lys

PCT/US97/22578

	210		215		220															
5	Ile Ile Ile Pro Phe Gly Leu Ala Leu Leu Leu Glu Ala Leu Leu Leu 225           230	Lys Gly Leu Pro Leu Lys Ser Ser Val Val Asn Ser Ser Thr Ala Leu 245                 250                 255	Leu Gly Met Leu Leu Pro Lys Gly Ile Ala Leu Leu Thr Ile Thr Ser Leu 260                 265                 270	Leu Thr Ala Val Ile Lys Leu Gly Leu Lys Lys Val Leu Val Gln Glu 275                 280                 285	Met Tyr Ser Val Glu Thr Leu Ala Arg Val Asp Met Leu Cys Leu Asp 290                 295                 300	Lys Thr Gly Thr Ile Thr Gln Gly Lys Met Gln Val Glu Ala Val Leu 305                 310                 315	Pro Leu Thr Glu Thr Tyr Gly Glu Glu Ala Ile Ala Ser Ile Leu Thr 325                 330                 335	Ser Tyr Met Ala His Ser Glu Asp Lys Asn Pro Thr Ala Gln Ala Ile 340                 345                 350	Arg Gln Arg Leu Trp Glu Met Leu Leu Ile 355                 360											
30	(ii) INFORMATION FOR SEQ ID NO:202:																			
	(i) SEQUENCE CHARACTERISTICS:																			
	(A) LENGTH: 243 amino acids																			
35	(B) TYPE: amino acid																			
	(C) STRANDEDNESS: not relevant																			
	(D) TOPOLOGY: not relevant																			
	(iii) MOLECULE TYPE: peptide																			
40	(iv) HYPOTHETICAL: NO																			
	(v) ANTI-SENSE: NO																			
45	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:202:																			
	Ala Ser Asn Ile Met Phe Met Leu Asp Leu Gly Asn His Leu Asp Gln 1                      5                      10                      15	Trp Ser Leu Lys Lys Thr Ala Thr Asn Leu Glu Gln Ser Leu Leu Ala 20                      25                      30	Lys Glu Ser Asp Val Phe Leu Val Gln Gly Asp Thr Val Val Ser Ile 35                      40                      45	Lys Ser Ser Asp Val Gln Ile Gly Asp Val Leu Ile Leu Ser Gln Gly 50                      55                      60	Asn Glu Ile Leu Phe Asp Gly Gln Val Val Ser Gly Leu Gly Met Val 65                      70                      75                      80															

PCT/US97/22578

[illegible]

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 307 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:203:

	Lys	Gln	Ile	Glu	Val	Val	Asp	Lys	Asp	Asn	Lys	Ser	Glu	Thr	Ala	Glu	
	1			5					10						15		
55	Ala	Ala	Ser	Val	Thr	Thr	Asn	Leu	Val	Thr	Gln	Ser	Lys	Val	Ser	Ala	
				20				25						30			
	Val	Val	Gly	Pro	Ala	Thr	Ser	Gly	Ala	Thr	Ala	Ala	Ala	Val	Ala	Asn	
			35					40					45				
60	Ala	Thr	Lys	Ala	Gly	Val	Pro	Leu	Ile	Ser	Pro	Ser	Ala	Thr	Gln	Asp	
	50						55					60					

WO 98/26072

PCT/US97/22578

-296-

Gly Leu Thr Lys Gly Gln Asp Tyr Leu Phe Ile Gly Thr Phe Gln Asp  
 65 70 75 80  
 5 Ser Phe Gln Gly Lys Ile Ile Ser Asn Tyr Val Ser Glu Lys Leu Asn  
 85 90 95  
 Ala Lys Lys Val Val Leu Tyr Thr Asp Asn Ala Ser Asp Tyr Ala Lys  
 100 105 110  
 10 Gly Ile Ala Lys Ser Phe Arg Glu Ser Tyr Lys Gly Glu Ile Val Ala  
 115 120 125  
 Asp Glu Thr Phe Val Ala Gly Asp Thr Asp Phe Gln Ala Ala Leu Thr  
 130 135 140  
 15 Lys Met Lys Gly Lys Asp Phe Asp Ala Ile Val Val Pro Gly Tyr Tyr  
 145 150 155 160  
 20 Asn Glu Ala Gly Lys Ile Val Asn Gln Ala Arg Gly Met Gly Ile Asp  
 165 170 175  
 Lys Pro Ile Val Gly Gly Asp Gly Phe Asn Gly Glu Glu Phe Val Gln  
 180 185 190  
 25 Gln Ala Thr Ala Glu Lys Ala Ser Asn Ile Tyr Phe Ile Ser Gly Phe  
 195 200 205  
 30 Ser Thr Thr Val Glu Val Ser Ala Lys Ala Lys Ala Phe Leu Asp Ala  
 210 215 220  
 Tyr Arg Ala Lys Tyr Asn Glu Glu Pro Ser Thr Phe Ala Ala Leu Ala  
 225 230 235 240  
 35 Tyr Asp Ser Val His Leu Val Ala Asn Ala Ala Lys Gly Ala Lys Asn  
 245 250 255  
 Ser Gly Glu Ile Lys Asn Asn Leu Ala Lys Thr Lys Asp Phe Glu Gly  
 260 265 270  
 40 Val Thr Gly Gln Thr Ser Phe Asp Ala Asp His Asn Thr Val Lys Thr  
 275 280 285  
 45 Ala Tyr Met Met Thr Met Asn Asn Gly Lys Val Glu Ala Ala Glu Val  
 290 295 300  
 Val Lys Pro  
 305

50 (2) INFORMATION FOR SEQ ID NO:204:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 289 amino acids  
 (B) TYPE: amino acid  
 55 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant  
 (ii) MOLECULE TYPE: peptide  
 60 (iii) HYPOTHETICAL: NO



WO 98/26072

PCT/US97/22578

-298-

(2) INFORMATION FOR SEQ ID NO:205:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 106 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

Ser	Gln	Asp	Gln	Thr	Trp	Tyr	Ala	Leu	Ala	Tyr	Asp	Gly	Ala	Glu	Val
1				5					10				15		
Ile	Gly	Phe	Leu	Thr	Val	Gln	Glu	Thr	Leu	Phe	Glu	Ala	Glu	Val	Leu
		20					25					30			
Gln	Ile	Ala	Val	Lys	Gly	Ala	Tyr	Gln	Gly	Gln	Gly	Ile	Ala	Ser	Ala
	35				40						45				
Leu	Phe	Ala	Gln	Leu	Pro	Thr	Asp	Lys	Glu	Ile	Phe	Leu	Glu	Val	Arg
	50				55				60						
Gln	Ser	Asn	Gln	Arg	Ala	Gln	Ala	Phe	Tyr	Lys	Lys	Glu	Lys	Met	Ala
	65			70					75					80	
Val	Ile	Ala	Glu	Arg	Lys	Ala	Tyr	Tyr	His	Asp	Pro	Val	Glu	Asp	Ala
			85						90					95	
Ile	Ile	Met	Lys	Arg	Glu	Ile	Asp	Glu	Gly						
			100				105								

(2) INFORMATION FOR SEQ ID NO:206:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 45 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:206:

Lys	Thr	Leu	Lys	Gly	His	Gly	Gln	Phe	Leu	His	Ala	Lys	Thr	Leu	Gly
1				5					10				15		
Phe	Thr	His	Pro	Arg	Thr	Gly	Lys	Thr	Leu	Glu	Phe	Lys	Ala	Asp	Ile
	20						25					30			



WO 98/26072

PCT/US97/22578

-299-

Pro Glu Ile Phe Lys Glu Thr Leu Glu Arg Leu Arg Lys  
35 40 45

5 (2) INFORMATION FOR SEQ ID NO:207:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 163 amino acids

(B) TYPE: amino acid

10 (C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

15 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

Arg Glu Met Val Val His Pro Ser Ala Gly His Thr Ser Gly Thr Leu  
1 5 10 15

25 Val Asn Ala Leu Met Tyr His Ile Lys Asp Leu Ser Gly Ile Asn Gly  
20 25 30

Val Leu Arg Pro Gly Ile Val His Arg Ile Asp Lys Asp Thr Ser Gly  
35 40 45

30 Leu Leu Met Ile Ala Lys Asn Asp Asp Ala His Leu Val Leu Ala Gln  
50 55 60

35 Glu Leu Lys Asp Lys Lys Ser Leu Arg Lys Tyr Trp Ala Ile Val His  
65 70 75 80

Gly Asn Leu Pro Asn Asp Arg Gly Val Ile Glu Ala Pro Ile Gly Arg  
85 90 95

40 Ser Glu Lys Asp Arg Lys Lys Gln Ala Val Thr Ala Lys Gly Lys Pro  
100 105 110

Ala Val Thr Arg Phe His Val Leu Glu Arg Phe Gly Asp Tyr Ser Leu  
115 120 125

45 Val Glu Leu Gln Leu Glu Thr Gly Arg Thr His Gln Ile Arg Val His  
130 135 140

50 Met Ala Tyr Ile Gly His Pro Val Ala Gly Asp Glu Val Tyr Gly Pro  
145 150 155 160

Ala Arg Leu

55 (2) INFORMATION FOR SEQ ID NO:208:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 224 amino acids

(B) TYPE: amino acid

60 (C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

WO 98/26072

PCT/US97/22578

-300-

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

Leu Gly Thr Arg Gly Ser Ser Arg Val Asp Asn Ile Asn Leu Gln Val  
1 5 10 15

Asp Glu Arg Asp Arg Ile Ala Leu Val Gly Lys Asn Gly Ala Gly Lys  
20 25 30

Ser Thr Leu Leu Lys Ile Leu Val Gly Glu Glu Glu Pro Thr Ser Gly  
35 40 45

Glu Ile Asn Lys Lys Lys Asp Ile Ser Leu Ser Tyr Leu Ala Gln Asp  
50 55 60

Ser Arg Phe Glu Ser Glu Asn Thr Ile Tyr Asp Glu Met Leu His Val  
65 70 75 80

Phe Asn Asp Leu Arg Arg Thr Glu Arg Gln Leu Arg Gln Met Glu Leu  
85 90 95

Glu Met Gly Glu Lys Ser Gly Glu Asp Leu Asp Lys Leu Met Ser Asp  
100 105 110

Tyr Asp Arg Leu Ser Glu Asn Phe Arg Gln Ala Gly Gly Phe Thr Tyr  
115 120 125

Glu Ala Asp Ile Arg Ala Ile Leu Asn Gly Phe Lys Phe Asp Glu Ser  
130 135 140

Met Trp Gln Met Lys Ile Ala Glu Leu Ser Gly Gly Gln Asn Thr Arg  
145 150 155 160

Leu Ala Leu Ala Lys Met Leu Leu Glu Lys Pro Asn Leu Leu Val Leu  
165 170 175

Asp Glu Pro Thr Asn His Leu Asp Ile Glu Thr Ile Ala Trp Leu Glu  
180 185 190

Asn Tyr Leu Val Asn Tyr Ser Gly Ala Leu Ile Ile Val Ser His Asp  
195 200 205

Arg Tyr Phe Leu Asp Lys Val Ala Thr Ile Thr Leu Asp Leu Thr Ser  
210 215 220

(2) INFORMATION FOR SEQ ID NO:209:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

-301-

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

5 (iv) ANTI-SENSE: NO

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:

Ser Thr Thr His His Leu Leu Val Lys Lys Val Asn Gly Leu Leu Val  
1 5 10 15

15 Arg Trp Lys Asn Ala Cys Arg Gln Asn Cys Lys Gln Thr Phe Xaa Phe  
20 25 30

Val Leu Thr Gln Leu Ile His Ala Asp Lys Trp Thr Val Ser Gly Arg  
35 40 45

20 Gly Glu Leu His Leu Ser Ile Leu Ile Glu Thr Met Arg Arg Glu Gly  
50 55 60

25 Tyr Glu Leu Gln Val Ser Arg Pro Glu Val Ile Val Lys Glu Ile Asp  
65 70 75 80

Gly Val Lys Cys Glu Pro Phe Glu Arg Val Gln Ile Asp Thr Pro Glu  
85 90 95

30 Glu Tyr Gln Gly Ser Val Ile Gln Ser Leu Ser Glu Arg Lys Gly Glu  
100 105 110

Met Leu Asp Met Ile Ser Thr Gly Asn Gly Gln Thr Arg Leu Val Phe  
115 120 125

35 Leu Val Pro Ala Arg Gly Leu Xaa Trp Ile Leu Asn Val Leu Val Asn  
130 135 140

40 Asp Ser Trp Leu Arg Tyr His Glu Pro Tyr Leu Arg Pro Ile Leu Ala  
145 150 155 160

Ile Asp Ser Arg Gly Asn Trp Trp Thr Ser Pro Trp Cys Pro Cys Phe  
165 170 175

45 Tyr Arg Cys Trp Gly Tyr Asn Leu Leu Asn Leu Leu Ser Thr Leu  
180 185 190

50 (2) INFORMATION FOR SEQ ID NO:210:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 252 amino acids

(B) TYPE: amino acid

55 (C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

60 (iv) ANTI-SENSE: NO

WO 98/26072

PCT/US97/22578

-302-

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:210:

5	Met	Phe	Gly	Phe	Phe	Lys	Lys	Asp	Lys	Ala	Val	Glu	Val	Glu	Val	Pro	1	5	10	15	20
10	Thr	Gln	Val	Pro	Ala	His	Ile	Gly	Ile	Ile	Met	Asp	Gly	Asn	Gly	Arg	20	25	30	35	40
15	Trp	Ala	Lys	Lys	Arg	Met	Gln	Pro	Arg	Val	Phe	Gly	His	Lys	Ala	Gly	35	40	45	50	55
20	Met	Glu	Ala	Leu	Gln	Thr	Val	Thr	Lys	Ala	Ala	Asn	Lys	Leu	Gly	Val	50	55	60	65	70
25	Lys	Val	Ile	Thr	Val	Tyr	Ala	Phe	Ser	Thr	Glu	Asn	Trp	Thr	Arg	Pro	70	75	80	85	90
30	Asp	Gln	Glu	Val	Lys	Phe	Xaa	Met	Asn	Leu	Pro	Val	Glu	Phe	Tyr	Asp	85	90	95	100	105
35	Asn	Tyr	Val	Pro	Glu	Leu	His	Ala	Asn	Asn	Val	Lys	Ile	Gln	Met	Ile	100	105	110	115	120
40	Gly	Glu	Thr	Asp	Arg	Leu	Pro	Lys	Gln	Thr	Phe	Glu	Ala	Leu	Thr	Lys	115	120	125	130	135
45	Ala	Glu	Glu	Leu	Thr	Lys	Asn	Asn	Thr	Gly	Leu	Ile	Leu	Asn	Phe	Ala	130	135	140	145	150
50	Leu	Asn	Tyr	Gly	Gly	Arg	Ala	Glu	Ile	Thr	Gln	Ala	Leu	Lys	Leu	Ile	145	150	155	160	165
55	Ser	Gln	Asp	Val	Leu	Asp	Ala	Lys	Ile	Asn	Pro	Gly	Asp	Ile	Thr	Glu	165	170	175	180	185
60	Glu	Leu	Ile	Gly	Asn	Tyr	Leu	Phe	Thr	Gln	His	Leu	Pro	Lys	Asp	Leu	180	185	190	195	200
65	Arg	Asp	Pro	Asp	Leu	Ile	Ile	Arg	Thr	Ser	Gly	Glu	Leu	Arg	Leu	Ser	195	200	205	210	215
70	Asn	Phe	Leu	Pro	Trp	Gln	Gly	Ala	Tyr	Ser	Glu	Leu	Tyr	Phe	Thr	Asp	210	215	220	225	230
75	Thr	Leu	Trp	Pro	Asp	Phe	Asp	Glu	Ala	Ala	Leu	Gln	Glu	Ala	Ile	Leu	225	230	235	240	245
80	Ala	Tyr	Asn	Arg	Arg	His	Arg	Arg	Phe	Gly	Gly	Val					245	250			

(2) INFORMATION FOR SEQ ID NO:211:

55 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 67 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant  
 60 (ii) MOLECULE TYPE: peptide

-303-

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:211:

Val Glu Gln Lys Leu Arg Gly Arg Asn Glu Asn Glu Ile Gln Ser Glu  
 1 5 10 15  
 Asp Ile Gly Ser Leu Val Met Glu Glu Leu Ala Glu Leu Asp Glu Ile  
 20 25 30  
 Thr Tyr Val Arg Phe Ala Ser Val Tyr Arg Ser Phe Lys Asp Val Ser  
 35 40 45  
 Glu Leu Glu Ser Leu Leu Gln Gln Ile Thr Gln Ser Ser Lys Lys Lys  
 50 55 60  
 Lys Glu Arg  
 65

(2) INFORMATION FOR SEQ ID NO:212:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 75 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:212:

Val Asp Ser Arg Gln Ala Glu Glu Gly Asn Thr Ile Arg Arg Arg Arg  
 1 5 10 15  
 Glu Cys Asp Glu Cys Gln His Arg Phe Thr Thr Tyr Glu Arg Val Glu  
 20 25 30  
 Glu Arg Thr Leu Val Val Val Lys Lys Asp Gly Thr Arg Glu Gln Phe  
 35 40 45  
 Ser Arg Asp Lys Ile Phe Asn Gly Ile Ile Arg Ser Ala Gln Lys Arg  
 50 55 60  
 Pro Val Ser Ser Asp Glu Ile Asn Met Val Ile  
 65 70 75

(2) INFORMATION FOR SEQ ID NO:213:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 63 amino acids

WO 98/26072

PCT/US97/22578

-304-

(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

- 5 (ii) MOLECULE TYPE: peptide  
(iii) HYPOTHETICAL: NO  
10 (iv) ANTI-SENSE: NO  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:213:
- 15 Phe Ala Gln Val Pro Lys Val Ala Gln Lys Val Met Lys Val Thr Lys  
1 5 10 15  
Ala Ala Gly Met Asn Ile Ile Ser Asn Cys Glu Glu Val Ala Gly Gln  
20 20 25 30  
Thr Val Phe His Thr His Val His Leu Val Pro Arg Tyr Ser Ala Asp  
35 40 45  
Asp Asp Leu Lys Ile Asp Phe Ile Ala His Glu Thr Asp Phe Asp  
50 55 60
- 25 (2) INFORMATION FOR SEQ ID NO:214:
- (i) SEQUENCE CHARACTERISTICS:  
30 (A) LENGTH: 67 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide  
35 (iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: NO  
40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:214:
- Met Ser Asp Cys Ile Phe Cys Lys Ile Ile Ala Gly Glu Ile Pro Ala  
45 1 5 10 15  
Ser Lys Val Tyr Glu Asp Glu Gln Val Leu Ala Phe Leu Asp Ile Ser  
20 25 30  
Gln Val Thr Leu Gly His Thr Leu Val Val Pro Lys Glu His Tyr Arg  
50 35 40 45  
Asn Leu Leu Glu Met Asp Ala Thr Ser Ala Thr Asn Ser Leu Pro Lys  
55 50 55 60  
Tyr Gln Lys  
65
- (2) INFORMATION FOR SEQ ID NO:215:
- 60 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 212 amino acids



WO 98/26072

PCT/US97/22578

-306-

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

5 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:

Leu Leu Ser Leu Ile Asp Ile Leu Val Asp Gly Arg Tyr Asp Arg Thr  
1 5 10 15

15 Lys Arg Asn Leu Met Leu Gln Phe Arg Gly Ser Ser Asn Gln Arg Ile  
20 25 30

Ile Asp Ser Arg Gly Ser Pro Gly Thr Glu Leu  
35 40

20 (2) INFORMATION FOR SEQ ID NO:217:

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 130 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

30 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:217:

Met Asn Asn Pro Lys Pro Gln Glu Trp Lys Ser Glu Glu Leu Ser Gln  
1 5 10 15

40 Gly Arg Ile Ile Asp Tyr Lys Ala Phe Asn Phe Val Asp Gly Glu Gly  
20 25 30

45 Val Arg Asn Ser Leu Tyr Val Ser Gly Cys Met Phe His Cys Glu Gly  
35 40 45

Cys Tyr Asn Val Ala Thr Trp Ser Phe Asn Ala Gly Ile Pro Tyr Thr  
50 55 60

50 Ala Glu Leu Glu Glu Gln Ile Met Ala Asp Leu Ala Gln Pro Tyr Val  
65 70 75 80

Gln Gly Leu Thr Leu Leu Gly Gly Glu Pro Phe Leu Asn Thr Gly Ile  
85 90 95

55 Leu Leu Pro Leu Val Lys Arg Ile Arg Lys Glu Leu Pro Asp Lys Asp  
100 105 110

60 Ile Trp Ser Trp Thr Gly Tyr Thr Trp Glu Glu Met Ile Pro Gly Asn  
115 120 125



PCT/US97/22578

Ser Arg  
130

5

10

15

(iv) ANTI-SENSE: NO

20

25

30

35

40

Leu Asp Lys Thr Lys Glu Glu Gly Val Ala Pro Leu Phe Trp  
115 120 125

45

50

55

(iv) ANTI-SENSE: NO

60



WO 98/26072

PCT/US97/22578

-309-

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:

Leu Met His Val Thr Val Gly Glu Leu Ile Gly Asn Phe Ile Leu Ile  
1 5 10 15  
10 Thr Gly Ser Phe Ile Leu Leu Leu Val Leu Ile Lys Lys Phe Ala Trp  
20 25 30  
15 Ser Asn Ile Thr Gly Ile Phe Glu Glu Arg Ala Glu Lys Ile Ala Ser  
35 40 45  
Asp Ile Asp Arg Ala Glu Glu Ala Arg Gln Lys Ala Val Leu Ala  
50 55 60  
20 Gln Lys Arg Glu Asp Glu Leu Ala Gly Ser Arg Lys Glu Ala Lys Thr  
65 70 75 80  
Ile Ile Glu Asn Ala Lys Glu Thr Ala Glu Gln Ser Lys Ala Asn Ile  
85 90 95  
25 Leu Ala Asp Ala Lys Leu Glu Ala Gly His Leu Lys Glu Lys Ala Asn  
100 105 110  
30 Gln Glu Ile Ala Gln Asn Lys Val Glu Ala Leu Gln Ser Val Lys Gly  
115 120 125  
Glu Val Ala Asp Leu Thr Ile Ser Leu Ala Gly Lys Ile Ile Ser Gln  
130 135 140  
35 Asn Leu Asp Ser His Ala His Lys Ala Leu Ile Asp Gln Tyr Ile Asp  
145 150 155 160  
Gln Leu Gly Glu Ala  
165  
40

(2) INFORMATION FOR SEQ ID NO:221:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 629 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

45

(ii) MOLECULE TYPE: peptide

50

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:221:

Met Gln Arg Leu Val Ser Leu Leu Ile Trp Ser Leu Leu Glu Thr Ser  
1 5 10 15  
60 Ile Leu Ser Ile His Gly Leu Gly Pro Leu Thr Lys Arg Phe Gly Val



WO 98/26072

PCT/US97/22578

-311-

Ser Gln Met Leu Val Gln Arg Ser Asn Glu Arg Gly Tyr Leu Val Gly  
 355 360 365  
 5 Ser Arg Gly Ser Val Gly Ser Ser Phe Val Ala Thr Met Ile Gly Ile  
 370 375 380  
 Thr Glu Val Asn Pro Leu Ser Pro His Tyr Val Cys Gly Gln Cys Gln  
 385 390 395 400  
 10 Tyr Ser Glu Phe Ile Thr Asp Gly Ser Tyr Gly Ser Gly Phe Asp Met  
 405 410 415  
 15 Pro His Lys Asp Cys Pro Asn Cys Gly His Lys Leu Ser Lys Asn Gly  
 420 425 430  
 Gln Asp Ile Pro Phe Glu Thr Phe Leu Gly Phe Asp Gly Asp Lys Val  
 435 440 445  
 20 Pro Asp Ile Asp Leu Asn Phe Ser Ser Gly Glu Asp Gln Pro Ser Ala His  
 450 455 460  
 25 Leu Asp Val Arg Asp Ile Phe Gly Glu Glu Tyr Ala Phe Arg Ala Gly  
 465 470 475 480  
 Thr Val Gly Thr Val Ala Ala Lys Thr Ala Tyr Gly Phe Val Lys Gly  
 485 490 495  
 30 Tyr Glu Arg Asp Tyr Gly Lys Phe Tyr Arg Asp Ala Glu Val Glu Arg  
 500 505 510  
 Leu Ala Gln Gly Ala Ala Gly Val Lys Arg Thr Thr Gly Gln His Pro  
 515 520 525  
 35 Gly Gly Ile Val Val Ile Pro Asn Tyr Met Asp Val Tyr Asp Phe Thr  
 530 535 540  
 40 Pro Val Gln Tyr Pro Ala Asp Asp Val Thr Ala Glu Trp Gln Thr Thr  
 545 550 555 560  
 His Phe Asn Phe His Asp Ile Asp Glu Asn Val Leu Lys Leu Asp Val  
 565 570 575  
 45 Leu Gly His Asp Asp Pro Thr Met Ile Arg Lys Leu Gln Asp Leu Ser  
 580 585 590  
 Gly Ile Asp Pro Asn Lys Ile Pro Met Asp Asp Glu Gly Val Met Ala  
 595 600 605  
 50 Leu Phe Ser Gly Thr Asp Val Leu Gly Val Thr Pro Glu Gln Ile Gly  
 610 615 620  
 Thr Leu Arg Val Cys  
 625

55

(2) INFORMATION FOR SEQ ID NO:222:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 693 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant

60

WO 98/26072

PCT/US97/22578

-312-

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

5 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:

15 Met Ala Arg Glu Phe Ser Leu Glu Lys Thr Arg Asn Ile Gly Ile Met  
1 5 10  
Ala His Val Asp Ala Gly Lys Thr Thr Thr Thr Glu Arg Ile Leu Tyr  
20 20 25 30  
Tyr Thr Gly Lys Ile His Lys Ile Gly Glu Thr His Glu Gly Ala Ser  
35 40 45  
Gln Met Asp Trp Met Glu Gln Glu Gln Glu Arg Gly Ile Thr Ile Thr  
25 50 55 60  
Ser Ala Ala Thr Thr Ala Gln Trp Asn Asn His Arg Val Asn Ile Ile  
65 70 75 80  
Asp Thr Pro Gly His Val Asp Phe Thr Ile Glu Val Gln Arg Ser Leu  
30 85 90 95  
Arg Val Leu Asp Gly Ala Val Thr Val Leu Asp Ser Gln Ser Gly Val  
100 105 110  
Glu Pro Gln Thr Glu Thr Val Trp Arg Gln Ala Thr Glu Tyr Gly Val  
35 115 120 125  
Pro Arg Ile Val Phe Ala Asn Lys Met Asp Lys Ile Gly Ala Asp Phe  
40 130 135 140  
Leu Tyr Ser Val Ser Thr Leu His Asp Arg Leu Gln Ala Asn Ala His  
145 150 155 160  
Pro Ile Gln Leu Pro Ile Gly Ser Glu Asp Asp Phe Arg Gly Ile Ile  
45 165 170 175  
Asp Leu Ile Lys Met Lys Ala Glu Ile Tyr Thr Asn Asp Leu Gly Thr  
180 185 190  
Asp Ile Leu Glu Glu Asp Ile Pro Ala Glu Tyr Leu Asp Gln Ala Gln  
50 195 200 205  
Glu Tyr Arg Glu Lys Leu Ile Glu Ala Val Ala Glu Thr Asp Glu Glu  
55 210 215 220  
Leu Met Met Lys Tyr Leu Glu Gly Glu Glu Ile Thr Asn Glu Glu Leu  
225 230 235 240  
Lys Ala Gly Ile Arg Lys Ala Thr Ile Asn Val Glu Phe Phe Pro Val  
60 245 250 255

WO 98/26072

PCT/US97/22578

-313-

Leu Cys Gly Ser Ala Phe Lys Asn Lys Gly Val Gln Leu Met Leu Asp  
 260 265  
 5 Ala Val Ile Asp Tyr Leu Pro Ser Pro Leu Asp Ile Pro Ala Ile Lys  
 275 280 285  
 Gly Ile Asn Pro Asp Thr Asp Ala Glu Glu Ile Arg Pro Ala Ser Asp  
 290 295 300  
 10 Glu Glu Pro Phe Ala Ala Leu Ala Phe Lys Ile Met Thr Asp Pro Phe  
 305 310 315 320  
 Val Gly Arg Leu Thr Phe Phe Arg Val Tyr Ser Gly Val Leu Gln Ser  
 325 330 335  
 15 Gly Ser Tyr Val Leu Asn Thr Ser Lys Gly Lys Arg Glu Arg Ile Gly  
 340 345 350  
 20 Arg Ile Leu Gln Met His Ala Asn Ser Arg Gln Glu Ile Asp Thr Val  
 355 360 365  
 Tyr Ser Gly Asp Ile Ala Ala Val Gly Leu Lys Asp Thr Thr Thr  
 370 375 380  
 25 Gly Asp Ser Leu Thr Asp Glu Lys Ala Lys Ile Ile Leu Glu Ser Ile  
 385 390 395 400  
 Asn Val Pro Glu Pro Val Ile Gln Leu Met Val Glu Pro Lys Ser Lys  
 405 410 415  
 30 Ala Asp Gln Asp Lys Met Gly Ile Ala Leu Gln Lys Leu Ala Glu Glu  
 420 425 430  
 35 Asp Pro Thr Phe Arg Val Glu Thr Asn Val Glu Thr Gly Glu Thr Val  
 435 440 445  
 Ile Ser Gly Met Gly Glu Leu His Leu Asp Val Leu Val Asp Arg Met  
 450 455 460  
 40 Arg Arg Glu Phe Lys Val Glu Ala Asn Val Gly Ala Pro Gln Val Ser  
 465 470 475 480  
 Tyr Arg Glu Thr Phe Arg Ala Ser Thr Gln Ala Arg Gly Phe Phe Lys  
 485 490 495  
 45 Arg Gln Ser Gly Gly Lys Gly Gln Phe Gly Asp Val Trp Ile Glu Phe  
 500 505 510  
 50 Thr Pro Asn Glu Glu Gly Lys Gly Phe Glu Phe Glu Asn Ala Ile Val  
 515 520 525  
 Gly Gly Val Val Pro Arg Glu Phe Ile Pro Ala Val Glu Lys Gly Leu  
 530 535 540  
 55 Val Glu Ser Met Ala Asn Gly Val Leu Ala Gly Tyr Pro Met Val Asp  
 545 550 555 560  
 Val Lys Ala Lys Leu Tyr Asp Gly Ser Tyr His Asp Val Asp Ser Ser  
 565 570 575  
 60 Glu Thr Ala Phe Lys Ile Ala Ala Ser Leu Ser Leu Lys Glu Ala Ala

WO 98/26072

PCT/US97/22578

-314-

		580		585		590
	Lys Ser Ala Gln Pro Ala Ile Leu Glu Pro Met Met Leu Val Thr Ile	595		600		605
5	Thr Val Pro Glu Glu Asn Leu Gly Asp Val Met Gly His Val Thr Ala	610		615		620
	Arg Arg Gly Arg Val Asp Gly Met Glu Ala His Gly Asn Ser Gln Ile	625		630		635
10	Val Arg Ala Tyr Val Pro Leu Ala Glu Met Phe Gly Tyr Ala Thr Val	645		650		655
	Leu Arg Ser Ala Ser Gln Gly Arg Gly Thr Phe Met Met Val Phe Asp	660		665		670
15	His Tyr Glu Asp Val Pro Lys Ser Val Gln Glu Glu Ile Ile Lys Lys	675		680		685
20	Asn Lys Gly Glu Asp	690				
25	(2) INFORMATION FOR SEQ ID NO:223:					
	(i) SEQUENCE CHARACTERISTICS:					
	(A) LENGTH: 274 amino acids					
	(B) TYPE: amino acid					
30	(C) STRANDEDNESS: not relevant					
	(D) TOPOLOGY: not relevant					
	(ii) MOLECULE TYPE: peptide					
35	(iii) HYPOTHETICAL: NO					
	(iv) ANTI-SENSE: NO					
40	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:223:					
	Ala Tyr Lys Gly His Gln Glu Tyr Val Leu Pro Gln Ala Ala Arg Lys	5		10		15
45	Ile Tyr Ala Tyr Arg Arg Tyr Asp Leu Asn Glu Ser Pro Lys Thr Ala	20		25		30
	Leu Asp Leu Ile Ile Pro Asp Leu Phe Leu His Ile Leu Asn Pro Ala	35		40		45
50	Glu Arg Glu Arg Lys Leu Lys Arg Glu Gly Val Glu Glu Leu Tyr Leu	50		55		60
	Leu Asp Phe Ser Ser Gln Phe Ala Ser Leu Thr Ala Gln Glu Phe Phe	65		70		75
55	Ala Thr Tyr Ile Lys Ala Met Asn Ala Lys Ile Ile Val Ala Gly Phe	85		90		95
60	Asp Tyr Thr Phe Gly Ser Asp Lys Lys Thr Ala Glu Asp Leu Lys Asp	100		105		110



PCT/US97/22578

[illegible]

## 35

- 40

- 45

- (iv) ANTI-SENSE: NO

- 50

55

60

Arg Ile Asn Leu Lys Asp Gly Phe Ala Gln Val Lys Asn Gly Glu Val  
50 55 60

WO 98/26072

PCT/US97/22578

-316-

Trp Leu Ser Asn Val His Ile Ala Pro Tyr Glu Glu Gly Asn Ile Trp  
65 70 75 80

5 Asn Gln Glu Pro Glu Arg Arg Arg Lys Leu Leu His Lys Lys Gln  
85 90 95

Ile Gln Lys Leu Glu Gln Glu Thr Lys Gly Thr Gly Met Thr Leu Val  
10 100 105 110

Pro Leu Lys Val Tyr Met Ala Thr Leu Ser Phe Phe  
115 120

(2) INFORMATION FOR SEQ ID NO:225:

15 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 441 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant  
(D) TOPOLOGY: not relevant

20 (ii) MOLECULE TYPE: peptide

25 (iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: NO

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:

Ile Val Lys Glu Glu Lys Gly Leu Lys Glu Lys Gln Phe Trp Asn Arg  
1 5 10 15

35 Ile Leu Glu Phe Ala Gln Glu Arg Leu Thr Arg Ser Met Tyr Asp Phe  
20 25 30

40 Tyr Ala Ile Gln Ala Glu Leu Ile Lys Val Glu Glu Asn Val Ala Thr  
35 40 45

Ile Phe Leu Pro Arg Ser Glu Met Glu Met Val Trp Glu Lys Gln Leu  
50 55 60

45 Lys Asp Ile Ile Val Val Ala Gly Phe Glu Ile Tyr Asp Ala Glu Ile  
65 70 75 80

Thr Pro His Tyr Ile Phe Thr Lys Pro Gln Asp Thr Thr Ser Ser Gln  
85 90 95

50 Val Glu Glu Ala Thr Asn Leu Thr Leu Tyr Asp Tyr Ser Pro Lys Leu  
100 105 110

55 Val Ser Ile Pro Tyr Ser Asp Thr Gly Leu Lys Glu Lys Tyr Thr Phe  
115 120 125

Asp Asn Phe Ile Gln Gly Asp Gly Asn Val Trp Ala Val Ser Ala Ala  
130 135 140

60 Leu Ala Val Ser Glu Asp Leu Ala Leu Thr Tyr Asn Pro Leu Phe Ile  
145 150 155 160

WO 98/26072

PCT/US97/22578

-317-

5 Tyr Gly Gly Pro Gly Leu Gly Lys Thr His Leu Leu Asn Ala Ile Gly  
165 170 175  
Asn Glu Ile Leu Lys Asn Ile Pro Asn Ala Arg Val Lys Tyr Ile Pro  
180 185 190  
10 Ala Glu Ser Phe Ile Asn Asp Phe Leu Asp His Leu Arg Leu Gly Glu  
195 200 205  
Met Glu Lys Phe Lys Lys Thr Tyr Arg Ser Leu Asp Leu Leu Ile  
210 215 220  
15 Asp Asp Ile Gln Ser Leu Ser Gly Lys Lys Val Ala Thr Gln Glu Glu  
225 230 235 240  
Phe Phe Asn Thr Phe Asn Ala Leu His Asp Lys Gln Lys Ile Val  
245 250 255  
20 Leu Thr Ser Asp Arg Ser Pro Lys His Leu Glu Gly Leu Glu Arg  
260 265 270  
Leu Val Thr Arg Phe Ser Trp Gly Leu Thr Gln Thr Ile Thr Pro Pro  
275 280 285  
25 Asp Phe Glu Thr Arg Ile Ala Ile Leu Gln Ser Lys Thr Glu His Leu  
290 295 300  
30 Gly Tyr Asn Phe Gln Ser Asp Thr Leu Glu Tyr Leu Ala Gly Gln Phe  
305 310 315 320  
Asp Ser Asn Val Arg Asp Leu Glu Gly Ala Ile Asn Asp Ile Thr Leu  
325 330 335  
35 Ile Ala Arg Val Lys Lys Ile Lys Asp Ile Thr Ile Asp Ile Ala Ala  
340 345 350  
Glu Ala Ile Arg Ala Arg Lys Gln Asp Val Ser Gln Met Leu Val Ile  
355 360 365  
40 Pro Ile Asp Lys Ile Gln Thr Glu Val Gly Asn Phe Tyr Gly Val Ser  
370 375 380  
45 Ile Lys Glu Met Lys Gly Ser Arg Arg Leu Gln Asn Ile Val Leu Ala  
385 390 395 400  
Arg Gln Val Ala Met Tyr Leu Ser Arg Glu Leu Thr Asp Asn Ser Leu  
405 410 415  
50 Pro Lys Ile Gly Lys Glu Leu Gly Glu Lys Ser Tyr His Ser His Ser  
420 425 430  
Cys Pro Cys Gln Asn Lys Ile Leu Asn  
435 440  
55

(2) INFORMATION FOR SEQ ID NO:226:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 201 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: not relevant

60

-318-

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

5 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:226:

15 Glu Leu Val Ser Thr Met Tyr Phe Arg Phe Asp Tyr Tyr Ser Gln Asn  
 1 5 10  
 Leu Gly Glu Ile Phe Ala Ile Gly Met Val Val Gly His Leu Arg Trp  
 20 20 25 30  
 Leu Ile Thr Gly Ala Leu Val Leu Tyr Ile Phe Ala Asp Arg Lys Leu  
 35 40 45  
 Ile Asn Thr Trp Asp Phe Leu Asp Ile Ala Ala Pro Ser Val Met Ile  
 50 55 60  
 25 Ala Gln Ser Leu Gly Arg Trp Gly Asn Phe Phe Asn Gln Glu Ala Tyr  
 65 70 75 80  
 Gly Ala Thr Val Asp Asn Leu Asp Tyr Leu Pro Gly Phe Ile Arg Asp  
 30 85 90 95  
 Gln Met Tyr Ile Glu Gly Ser Tyr Arg Gln Pro Thr Phe Leu Tyr Glu  
 100 105 110  
 35 Ser Leu Trp Asn Leu Leu Gly Phe Ala Leu Ile Leu Ile Phe Arg Arg  
 115 120 125  
 Lys Trp Lys Ser Leu Arg Arg Gly His Ile Thr Ala Phe Tyr Leu Ile  
 40 130 135 140  
 Trp Tyr Gly Phe Gly Arg Met Val Ile Glu Gly Met Arg Thr Asp Ser  
 145 150 155 160  
 45 Leu Met Phe Phe Gly Leu Arg Val Ser Gln Trp Leu Ser Val Val Leu  
 165 170 175  
 Ile Gly Leu Gly Ile Met Ile Val Ile Tyr Gln Asn Arg Lys Lys Ala  
 180 185 190  
 50 Pro Tyr Tyr Ile Thr Glu Glu Asn  
 195 200

(2) INFORMATION FOR SEQ ID NO:227:

55 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 491 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: not relevant  
 (D) TOPOLOGY: not relevant  
 60 (ii) MOLECULE TYPE: peptide

WO 98/26072

PCT/US97/22578

-319-

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:

```

10  Leu Glu Asp Phe Pro Leu Ser Val Thr Asn Pro Tyr Gly Arg Thr Lys
    1           5           10           15

    Leu Met Leu Glu Glu Ile Leu Thr Asp Ile Tyr Lys Ala Asp Ser Glu
    20           25           30

15  Trp Asn Val Val Leu Leu Arg Tyr Phe Asn Pro Ile Gly Val His Glu
    35           40           45

    Ser Gly Asp Leu Gly Glu Asn Pro Asn Gly Ile Pro Asn Asn Leu Leu
    50           55           60

20  Pro Tyr Val Thr Gln Val Ala Val Gly Lys Leu Glu Gln Val Gln Val
    65           70           75           80

25  Phe Gly Asp Asp Tyr Asp Thr Glu Asp Gly Thr Gly Val Arg Asp Tyr
    85           90           95

    Ile His Val Val Asp Leu Ala Lys Gly His Val Ala Ala Leu Lys Lys
    100          105          110

30  Ile Gln Lys Gly Ser Gly Leu Asn Val Tyr Asn Leu Gly Thr Gly Lys
    115          120          125

    Gly Tyr Ser Val Leu Glu Ile Ile Gln Asn Met Glu Lys Ala Val Gly
    130          135          140

35  Cys Pro Ile Pro Tyr Arg Ile Val Glu Arg Arg Pro Gly Asp Ile Ala
    145          150          155          160

40  Ala Cys Tyr Ser Asp Pro Ala Lys Ala Lys Ala Glu Leu Gly Trp Glu
    165          170          175

    Ala Glu Leu Asp Ile Thr Gln Met Cys Glu Gly His Gly Val Gly Arg
    180          185          190

45  Ala Ser Ile Gln Met Asp Leu Lys Thr Lys Met Met Ile Ser Ile Ile
    195          200          205

    Val Pro Cys Leu Asn Glu Glu Glu Val Leu Pro Leu Phe Tyr Gln Ala
    210          215          220

50  Leu Glu Ala Leu Leu Pro Asp Leu Glu Thr Glu Ile Glu Tyr Val Phe
    225          230          235          240

    Val Asp Asp Gly Ser Ser Asp Gly Thr Leu Glu Leu Leu Lys Ala Tyr
    245          250          255

55  Arg Glu Gln Asn Pro Ala Val His Tyr Ile Ser Phe Ser Arg Asn Phe
    260          265          270

60  Gly Lys Glu Ala Ala Leu Tyr Ala Gly Leu Gln Tyr Ala Thr Gly Asp
    275          280          285

```

WO 98/26072

PCT/US97/22578

-320-

	Leu	Val	Val	Val	Met	Asp	Ala	Asp	Leu	Gln	Asp	Pro	Pro	Ser	Met	Leu
	290						295					300				
5	Phe	Glu	Met	Lys	Asn	Val	Leu	Asp	Lys	Asn	Val	Asp	Leu	Asp	Cys	Val
	305					310					315					320
	Gly	Thr	Arg	Arg	Thr	Ser	Arg	Glu	Gly	Glu	Pro	Phe	Phe	Arg	Ser	Phe
10					325					330					335	
	Cys	Ala	Val	Leu	Phe	Tyr	Arg	Leu	Met	Gln	Lys	Ile	Ser	Pro	Val	Ala
				340					345					350		
15	Leu	Pro	Ser	Gly	Val	Arg	Asp	Phe	Arg	Met	Met	Arg	Arg	Ser	Val	Val
				355				360						365		
	Asp	Ala	Ile	Leu	Ser	Leu	Thr	Glu	Ser	Asn	Arg	Phe	Ser	Lys	Gly	Leu
				370				375				380				
20	Phe	Ala	Trp	Val	Gly	Phe	Lys	Thr	His	Tyr	Leu	Asp	Tyr	Pro	Asn	Val
	385					390					395					400
	Glu	Arg	Gln	Ala	Gly	Lys	Thr	Ser	Trp	Ser	Phe	Arg	Gln	Leu	Phe	Phe
25					405					410					415	
	Tyr	Ser	Ile	Glu	Gly	Ile	Val	Asn	Phe	Ser	Asp	Phe	Pro	Leu	Thr	Ile
				420					425					430		
30	Ala	Phe	Val	Ala	Gly	Leu	Leu	Ser	Cys	Phe	Leu	Ser	Leu	Leu	Met	Thr
				435				440					445			
	Phe	Phe	Val	Val	Val	Arg	Thr	Leu	Ile	Leu	Gly	Asn	Pro	Thr	Ser	Gly
				450			455					460				
35	Trp	Thr	Ser	Leu	Met	Ala	Val	Ile	Leu	Tyr	Leu	Gly	Gly	Ile	Gln	Leu
	465					470					475					480
	Leu	Thr	Ile	Gly	Ile	Leu	Gly	Lys	Tyr	Asn	Gln					
					485					490						

(2) INFORMATION FOR SEQ ID NO:228:

## (1) SEQUENCE CHARACTERISTICS

- SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 277 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: not relevant

## (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

60 Val Ile Ile Ile Asp Asp Asn Tyr Ser Asn Val Asn Leu Arg Asn Lys  
1 5 10 15  
Ile Ile His Gln Phe Gly Tyr Thr Asn His Arg Ile Lys Leu Ile Leu

WO 98/26072

PCT/US97/22578

-321-

	20	25	30
5	Ser Asn Glu Asp Leu Gly Ala Thr Asn Ala Arg Asn Ile Gly Ile Lys 35 40 45		
	Asn Ser Arg Gly Lys Tyr Ile Ser Phe Leu Asp Asp Asp Asp Glu Tyr 50 55 60		
10	Met Pro Asp Arg Ile Leu Lys Leu Met Ala Cys Phe Lys Lys Ser Arg 65 70 75 80		
	Met Lys Asn Leu Ala Leu Val Tyr Ser Tyr Gly Ile Ile Ile Tyr Pro 85 90 95		
15	Asn Gly Thr Arg Glu Glu Glu Lys Thr Asp Phe Val Gly Asn Pro Leu 100 105 110		
	Phe Val Gln Met Val His Asn Ile Ala Gly Thr Ser Phe Trp Leu Cys 115 120 125		
20	Lys Lys Glu Val Leu Glu Leu Ile Asn Gly Phe Glu Lys Ile Asp Ser 130 135 140		
	His Gln Asp Gly Val Val Leu Leu Lys Leu Leu Ala Gln Gly Tyr Gln 145 150 155 160		
25	Ile Asp Ile Val Arg Glu Phe Leu Val Asn Tyr Tyr Ala His Ser Lys 165 170 175		
30	Glu Asn Gly Ile Thr Gly Val Thr Gln Lys Thr Ile Asn Ala Asp Glu 180 185 190		
	Glu Tyr Tyr Asn Tyr Cys Arg Lys Tyr Phe Asn Leu Leu Ser Phe Asn 195 200 205		
35	Glu Arg Ile Leu Val Thr Lys Lys Tyr Tyr Ser Leu Asn Ile Lys Arg 210 215 220		
	Leu Leu Leu Ile Gly Asp Lys Cys Lys Ala Leu Lys Val Ile Lys Lys 225 230 235 240		
40	Ala Arg Glu Glu Lys Ile Phe Asn Glu Phe Leu Phe Leu Lys Tyr Met 245 250 255		
45	Leu Leu Tyr Arg Ser Phe Phe Tyr Cys Ile Tyr Asp Asn Tyr Val Gln 260 265 270		
	Leu Lys Phe Arg Lys 275		

WO 98/26072

PCT/US97/22578

-322-

## CLAIMS

1. An isolated nucleic acid compound comprising a sequence identical to or substantially identical to a  
5 sequence selected from the group consisting of SEQ ID NO:1 through SEQ ID NO:86.

2. An isolated nucleic acid compound comprising a sequence identical to or substantially identical to a  
10 sequence selected from the group consisting of SEQ ID NO:87, SEQ ID NO:89, SEQ ID NO:91, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:97, SEQ ID NO:99, SEQ ID NO:101, SEQ ID NO:103, SEQ ID NO:105, SEQ ID NO:107, SEQ ID NO:109, SEQ ID NO:111, SEQ ID NO:113, SEQ ID NO:115, SEQ ID NO:117, SEQ ID NO:119, and SEQ  
15 ID NO:121.

3. A substantially purified protein or fragment thereof from *S. pneumoniae* wherein said protein is selected from the group consisting of SEQ ID NO:88, SEQ ID NO:90, SEQ  
20 ID NO:92, SEQ ID NO:94, SEQ ID NO:96, SEQ ID NO:98, SEQ ID NO:100, SEQ ID NO:102, SEQ ID NO:104, SEQ ID NO:106, SEQ ID NO:108, SEQ ID NO:110, SEQ ID NO:112, SEQ ID NO:114, SEQ ID NO:116, SEQ ID NO:118, SEQ ID NO:120, SEQ ID NO:122, and SEQ ID NO:123 through SEQ ID NO:228.  
25

4. An isolated nucleic acid compound encoding any one of the proteins or fragments thereof of Claim 3.

5. A vector comprising any one of the nucleic acid  
30 compounds of claims 1, 2, or 4.

6. A recombinant host containing any one of the vectors of claim 5.



-323-

7. A substantially purified protein from *Streptococcus pneumoniae* as in Claim 3 wherein said protein is an external target protein selected from Table 1.

5           8. A substantially purified protein from *Streptococcus pneumoniae* as in Claim 3 wherein said protein is a hypothetical protein selected from Table 1.

9. A substantially purified protein from *Streptococcus pneumoniae* as in Claim 3 wherein said protein is a cell wall synthetic protein selected from Table 1.

10          10. A substantially purified protein from *Streptococcus pneumoniae* as in Claim 3 wherein said protein is a minimal gene set protein selected from Table 1.

11. A DNA chip having arrayed thereon any at least 15 base pair fragment of any one or more of the nucleic acids of claim 1.

20          12. A DNA chip having arrayed thereon any at least 15 base pair fragment of any one or more of the nucleic acids of claim 2.

25          13. A method for evaluating gene expression in *Streptococcus pneumoniae* comprising the step of incubating a DNA chip of claim 11 or Claim 12 with cDNA prepared from *Streptococcus pneumoniae* under conditions suitable for hybridization of complementary nucleic acid sequences.

30          14. A computer readable medium having recorded thereon any one or more of the nucleotide sequences of Claims 1 or Claim 2.

-324-

15. A method for identifying virulence genes in *S. pneumoniae*, comprising the steps of:

- a) preparing a DNA chip as in claim 11,
- b) preparing labeled cDNAs from
  - 5 i) *S. pneumoniae* cells recovered from an in vivo environment, and
  - ii) *S. pneumoniae* cells grown in vitro,
- c) hybridizing individually the cDNAs of steps (b) (i) and (b) (ii) to a chip of step (a); and
- 10 d) identifying a genomic DNA fragment or fragments on said chip that hybridize to the cDNAs of (b) (i) but not with the cDNAs of (b) (ii).

16. An antibody that selectively binds to a  
15 protein or peptide of Claim 3.

17. An antibody that selectively binds to an  
external target protein, or fragment thereof, identified in  
Table 1.

20

18. A DNA chip of Claim 11 or Claim 12 further comprising a layer of *S. pneumoniae* cells wherein said layer contacts with said nucleic acids.

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/22578

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC(6) : P16a6c Soc Extra Sheet. US CL : 536/23.7, 23.1; 530/350, 387.1; 435/320.1, 252.3, 6; 360/1 According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) U.S. : 536/23.7, 23.1; 530/350, 387.1; 435/320.1, 252.3, 6; 360/1 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) STN, APS, chip#, dna#, oligo? MPSRCH of nucleic acid and amino acid sequence databases of IntelliGenetics (SEQ ID NOs 1-228)		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	FUKUNAGA et al. Expression cloning of a receptor for murine granulocyte colony-stimulating factor. Cell. 20 April 1990, Vol. 61, pages 341-350, see entire document.	11-13 and 15
Y	GEETHA-HABIB et al. Glycosylation site binding protein, a component of oligosaccharyl transferase, is highly similar to three other 57 kd liminal proteins of the ER. Cell. 23 September 1988, Vol. 54, pages 1053-1060, see entire document.	11-13 and 15
Y	STRAUSS et al. Complete nucleotide sequence of the genomic RNA of Sindbis virus. Virology. February 1984, Vol. 133, No. 1, pages 92-110, see entire document.	11-13 and 15
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "B" earlier document published on or after the international filing date document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "C" document referring to an oral disclosure, use, exhibition or other means "D" document published prior to the international filing date but later than the priority date claimed "E" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principles or theory underlying the invention "F" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "G" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "H" document member of the same patent family		
Date of the actual completion of the international search 25 MARCH 1998		Date of mailing of the international search report 09 APR 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230		Authorized officer JAMES MARTINELL Telephone No. (703) 305-6136

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/22578

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WILSON et al. 2.2 Mb of contiguous nucleotide sequence from chromosome III of <i>C. elegans</i> . Nature. 03 March 1994, Vol. 368, No. 6466, pages 32-38, see entire document.	11-13 and 15
Y	RESTREPO et al. Antigen diversity in the bacterium <i>B. hermsii</i> through "somatic" mutations in rearranged vmp genes. Cell. 09 September 1994, Vol. 78, pages 867-876, see entire document.	11-13 and 15
Y	WHITEWAY et al. Dominant negative selection of heterologous genes: Isolation of <i>Candida albicans</i> genes that interfere with <i>Saccharomyces cerevisiae</i> mating factor-induced cell cycle arrest. Proc. Natl. Acad. Sci. USA. 15 October 1992, Vol. 89, No. 20, pages 9410-9414, see entire document.	11-13 and 15
Y	LACKS et al. Genetic basis of the complementary DpnI and DpnII restriction systems of <i>S. pneumoniae</i> : An intracellular cassette mechanism. Cell. 26 September 1986, Vol. 46, pages 993-1000, see entire document.	11-13 and 15
Y	STEENBERGH et al. Structure and expression of the human calcitonin/CGRP genes. FEBS Letters. 01 December 1986, Vol. 209, No. 1, pages 97-103, see entire document.	11-13 and 15
Y	SHAMBAUGH et al. The spliceosomal U small nuclear RNAs of <i>Ascaris lumbricoides</i> . Molecular and Biochemical Parasitology. April 1994, Vol. 64, No. 2, pages 349-352, see entire document.	11-13 and 15
Y	MAXWELL et al. A novel NF-kappaB p65 spliced transcript lacking exons 6 and 7 in a non-small cell lung carcinoma cell line. Gene. 12 December 1995, Vol. 166, No. 2, pages 339-340, see entire document.	11-13 and 15
Y	KONYECSNI et al. Broad-host-range plasmid and M13 bacteriophage-derived vectors for promoter analysis in <i>Escherichia coli</i> and <i>Pseudomonas aeruginosa</i> . Gene. 30 December 1988, Vol. 74, No. 2, pages 375-386, see entire document.	11-13 and 15
Y	ALLIKMETS et al. Cloning and organization of the abc and mdl genes of <i>Escherichia coli</i> : Relationship to eukaryotic multidrug resistance. Gene. 22 December 1993, Vol. 136, Nos. 1 and 2, pages 231-236, see entire document.	11-13 and 15
Y	SCHWEIZER. Two plasmids, X1918 and Z1918, for easy recovery of the xylE and lacZ reporter genes. Gene. 30 November 1993, Vol. 134, No. 1, pages 89-91, see entire document.	11-13 and 15

INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/22578

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KRAFT et al. Sequence of the complete P protein gene and part of the M protein gene form the histidine transport operon of <i>Escherichia coli</i> compared to that of <i>Salmonella typhimurium</i> . <i>Nucleic Acids Res.</i> 26 October 1987, Vol. 15, No. 20, page 8568, see entire document.	11-13 and 15
Y	SCHULER et al., Characterization of the human gene encoding LBR, an integral protein of the nuclear envelope inner membrane., <i>J. Biol. Chem.</i> 15 April 1994, Vol. 269, No. 15, pages 11312-11317, see entire document.	11-13 and 15
Y	LOTTENBERG et al. Cloning, sequence analysis, and expression in <i>Escherichia coli</i> of a streptococcal plasmin receptor. <i>J. Bacteriol.</i> August 1992, Vol. 174, No. 16, pages 5204-5210, see entire document.	11-13 and 15
Y	COPPOLA et al. Sequence and transcriptional activity of the <i>Escherichia coli</i> K-12 chromosome region between <i>rrnC</i> and <i>ilvGMEDA</i> . <i>Gene</i> . 02 January 1991, Vol. 97, No. 1, pages 21-27, see entire document.	11-13 and 15
Y	MUTO et al. Preferential use of A- and U-rich codons for <i>Mycoplasma capricolum</i> ribosomal proteins S8 and L6. <i>Nucleic Acids Res.</i> 1984, Vol. 12, No. 12, pages 8209-8217, see entire document.	11-13 and 15
Y	RUDNER et al. The <i>spo0K</i> locus of <i>Bacillus subtilis</i> is homologous to the oligopeptide permease locus is required for sporulation and competence. <i>J. Bacteriology</i> . February 1991, Vol. 173, No. 3, pages 1388-1398, see entire document.	11-13 and 15
Y	GREEN et al. Isolation and characterisation of a bovine cDNA encoding eukaryotic initiation factor 2 $\alpha$ . <i>Biochim. Biophys. Acta</i> . 08 October 1991, Vol. 1090, No. 2, pages 277-280, see entire document.	11-13 and 15
Y	LOGEMANN et al. Modes of expression and common structural features of the complete phenylalanine ammonia-lyase gene family in parsley. <i>Proc. Natl. acad. Sci. USA</i> . 20 June 1995, Vol. 92, No. 13, pages 5905-5909, see entire document.	11-13 and 15
Y	SHERWOOD et al. Characterization of HIR1 and HIR2, two genes required for regulation of histone gene transcription in <i>Saccharomyces cerevisiae</i> . <i>Molec. Cell. Biol.</i> January 1993, Vol. 13, No. 1, pages 28-38, see entire document.	11-13 and 15

INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/22578

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	GAGNON et al. Cloning, sequencing and expression in <i>Escherichia coli</i> of the <i>ptsI</i> gene encoding enzyme I of the phosphoenolpyruvate:sugar phosphotransferase transport system from <i>Streptococcus salivarius</i> . <i>Gene</i> . 02 November 1992, Vol. 121, No. 1, pages 71-78, see entire document.	11-13 and 15-17
Y	PUYET et al. Characterization of the <i>Streptococcus pneumoniae</i> maltosaccharide regulator <i>MalR</i> , a member of the <i>LacI-GalR</i> family of repressors displaying distinctive genetic features. <i>J. Biol. Chem.</i> 05 December 1993, Vol. 268, No. 34, pages 25402-25408, see entire document.	11-13 and 15
Y	CROSS et al. Purification of CpG islands using a methylated DNA binding column. <i>Nature Genetics</i> . March 1994, Vol. 6, pages 236-244, see entire document.	11-13 and 15
Y	SALUJA et al. The genetic basis of colony opacity in <i>Streptococcus pneumoniae</i> : evidence for the effect of box elements on the frequency of phenotypic variation. <i>Molecular Microbiology</i> . April 1995, Vol. 16, No. 2, pages 215-227, see entire document.	11-13 and 15-17
Y	ROTHNIE et al. Retrotransposon-like nature of <i>Tp1</i> elements: implications for the organisation of highly repetitive, hypermethylated DNA in the genome of <i>Physarum polycephalum</i> . <i>Nucleic Acids Res.</i> 25 January 1991, Vol. 19, No. 2, pages 279-286, see entire document.	11-13 and 15
Y	GITT et al. A strong sequence homology exists between the major RNA polymerase $\delta$ factors of <i>Bacillus subtilis</i> and <i>Escherichia coli</i> . <i>J. Biol. Chem.</i> 25 June 1985, Vol. 260, No. 12, pages 7178-7185, see entire document.	11-13 and 15
Y	LEUVEN et al. Structure of the gene ( <i>LRP1</i> ) coding for the human $\alpha 2$ -macroglobulin receptor lipoprotein receptor-related protein. <i>Genomics</i> . 01 November 1994, Vol. 24, No. 1, pages 78-89, see entire document.	11-13 and 15
Y	YOSHIOKA et al. Nucleotide sequence of the promoter-distal region of the <i>tra</i> operon of plasmid R100, including <i>traI</i> (DNA helicase I) and <i>traD</i> genes. <i>J. Mol. Biol.</i> 05 July 1990, Vol. 214, No. 1, pages 39-53, see entire document.	11-13 and 15
Y	HUI et al. Genetic transformation in <i>Streptococcus pneumoniae</i> : Nucleotide sequence analysis shows <i>comA</i> , a gene required for competence induction, to be a member of the bacterial ATP-dependent transport protein family. <i>J. Bacteriol.</i> January 1991, Vol. 173, No. 1, pages 372-381, see entire document.	11-13 and 15

Form PCT/ISA/210 (continuation of second sheet)(July 1992)\*

INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/22578

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	BARON et al. Co-regulation of two gene activities by tetracycline via a bidirectional promoter. Nucleic Acids Res. 11 September 1995, Vol. 23, No. 17, pages 3605-3606, see entire document.	11-13 and 15
Y	BORK et al. Exploring the Mycoplasma capricolum genome: a minimal cell reveals its physiology. Molecular Microbiology. June 1995, Vol. 16, No. 5, pages 955-967, see entire document.	11-13 and 15
Y	CHAMBERS et al. The pMTL nic <sup>+</sup> cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. Gene. 15 August 1988, Vol. 68, No. 1, pages 139-149, see entire document.	11-13 and 15
Y	FENOLL et al. Molecular basis of the optochin-sensitive phenotype of pneumococcus: characterization of the genes encoding the F <sub>0</sub> complex of the Streptococcus pneumoniae and Streptococcus oralis H <sup>-</sup> -ATPases. Molecular Microbiology, May 1994, Vol. 12, No. 4, pages 587-598, see entire document.	11-13 and 15-17
Y	STEGLITZ-MORSORDORF et al. Cloning, heterologous expression, and sequencing of the Proteus vulgaris glnAntrBC operon and implications of nitrogen control on heterologous urease expression. FEMS Microbiology Letters. 15 January 1993, Vol. 106, No. 2, pages 157-164, see entire document.	11-13 and 15
Y	DE WIT et al. Structure of the gene for the human myeloid IgA Fc receptor (CD89). J. Immunol. 01 August 1995, Vol. 155, No. 3, pages 1203-1209, see entire document.	11-13 and 15
Y	WEBSTER et al. Nucleotide sequence of the galactose gene cluster of Kluyveromyces lactis. Nucleic Acids Res. 25 August 1988, Vol. 16, No. 16, pages 8192-8194, see entire document.	11-13 and 15
Y	KIRSCH et al. Chemotactic methylesterase promotes adaptation to high concentrations of attractant in Bacillus subtilis. J. Biol. Chem. 05 September 1993, Vol. 268, No. 25, pages 18610-18616, see entire document.	11-13 and 15
Y	YAMASHITA et al. Molecular characterization of a Streptococcus mutans mutant altered in environmental stress responses. J. Bacteriol. October 1993, Vol. 175, No. 19, pages 6220-6228, see entire document.	11-13 and 15-17
Y	FRASER et al. The minimal gene complement of Mycoplasma genitalium. Science. 20 October 1995, Vol. 270, pages 397-403, see entire document.	11-13 and 15

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/22578

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	TOMINAGA et al. Upstream region of a genomic gene for human mitochondrial transcription factor 1. <i>Biochim. Biophys. Acta.</i> 15 June 1992, Vol. 1131, No. 2, pages 217-219, see entire document.	11-13 and 15
Y	CORTI et al. Molecular cloning of cDNAs encoding human carnitine acetyltransferase and mapping of the corresponding gene to chromosome 9q34.1. <i>Genomics.</i> 01 September 1994, Vol. 23, No. 1, pages 94-99, see entire document.	11-13 and 15
Y	LUSIS et al. Cloning and expression of apolipoprotein B, the major protein of low and very low density lipoproteins. <i>Proc. Natl. Acad. Sci. USA.</i> July 1985, Vol. 82, No. 14, pages 4597-4601, see entire document.	11-13 and 15
Y	KOHLBRECHER et al. Staphylococcal phosphoenolpyruvate-dependent phosphotransferase system: Molecular cloning and nucleotide sequence of the <i>Staphylococcus carnosus</i> ptsI gene and expression and complementation studies of the gene product. <i>J. Bacteriol.</i> April 1992, Vol. 174, No. 7, pages 2208-2214, see entire document.	16 and 17
Y	FLEISCHMANN et al. Whole-genome random sequencing and assembly of <i>Haemophilus influenzae</i> Rd. <i>Science.</i> 28 July 1995, Vol. 269, pages 496-512, see entire document.	11-13 and 15-17
Y	PEASE et al. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. <i>Proc. Natl. Acad. Sci. USA.</i> 24 May 1994, Vol. 91, No. 11, pages 5022-5026, see entire document.	11-13 and 15
X,P	CHENG et al. The com locus controls genetic transformation in <i>Streptococcus pneumoniae</i> . <i>Molecular Microbiology.</i> February 1997, Vol. 23, No. 4, pages 683-692, see entire document.	1, 4-6, and 14
Y	MORIARTY et al. Antibodies to peptides detect new hepatitis B antigen: Serological correlation with hepatocellular carcinoma. <i>Science.</i> 25 January 1985, Vol. 227, pages 429-433, see entire document.	16 and 17
X	WO 96/10647 A (FUSO PHARMACEUTICAL INDUSTRIES, LTD.) 11 April 1996, see entire document.	1, 4-6, and 14
X	ALLOING et al. The ami locus of the Gram-positive bacterium <i>Streptococcus pneumoniae</i> is similar to binding protein-dependent transport operons of Gram-negative bacteria. <i>Molecular Microbiology.</i> 1990, Vol. 4, No. 4, pages 633-644, see entire document.	1, 4-6, and 14
Y		16 and 17



